Capacity-achieving Feedback Scheme for Gaussian Finite-State Markov Channels with Channel State Information

Jialing Liu, Nicola Elia, and Sekhar Tatikonda

Abstract

In this paper, we propose a capacity-achieving communication scheme for a Gaussian finite-state Markov channel (FSMC) with noiseless output feedback and with channel state information, subject to an average channel input power constraint. This scheme is derived from a control theoretic perspective and is based on the connections between feedback communication and feedback control over an FSMC. It also considerably reduces encoder and decoder complexity; shortens coding delay; has a doubly exponential reliability function if the channel state behavior is "typical"; and nontrivially extends the Schalkwijk-Kailath coding scheme for an additive white Gaussian noise (AWGN) channel to an FSMC.

Index Terms

Feedback communication, control-oriented feedback communication schemes, feedback capacity, feedback stabilization, finite-state Markov channels, channel state information

I. INTRODUCTION

There have been many achievements in the study of Markov channels, in which the time-varying fading gains (typically referred to as channel states) are modelled as Markov chains; see e.g. [1]–[8] and references therein. [1] obtains the capacity of a finite-state Markov channel (FSMC) with independent and identically distributed (i.i.d.) inputs and with a known channel transition structure but without channel state information (CSI). [2] solves for the capacity of a Markov channel with instantaneous CSI at both the transmitter (or encoder) and receiver (or decoder), or at the receiver only. [3] investigates the capacity of a Markov channel with possibly imprecise or delayed CSI. [4] provides the capacity of an FSMC with CSI delayed at the transmitter side and instantaneous at the receiver side (DTRCSI). It also shows that, for a channel with DTRCSI, the access to the channel output by the transmitter via delayed feedback does not increase the capacity. [5] obtains the delay-constrained capacity for a flat block-fading channel with causal feedback. [8] studies the capacity of an FSMC with instantaneous, perfect CSI available at the receiver side at all time instants and at the transmitter side periodically every once in a fixed number of time instants. The Markov channels studied in above papers mainly focus on fading channels with white noise and without inter-symbol interference (ISI) when conditioned on CSI. More specifically, the channel state at each time is independent of the channel inputs up to that time if conditioned on past channel states, and the channel output at each time is independent of past channel inputs if conditioned on present channel state and input. Fading channels with ISI may also exhibit the Markov property and are closely related to the ISI-free Markov fading channels. For ISI channels with

J. Liu was with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA. He is now with Motorola Inc., Libertyville, IL 60048 USA (e-mail:jialingliu@motorola.com).

N. Elia is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: nelia@iastate.edu).

S. Tatikonda is with the Department of Electrical Engineering, Yale University, New Haven, CT 06520 USA (e-mail: sekhar.tatikonda@yale.edu).

output feedback, [6] characterizes the capacity and capacity-achieving distribution. [7] investigates the capacity of an FSMC with correlated inputs, with ISI, and with a known channel transition structure, but without CSI. Finally, we refer to a recent paper [9] for the study of capacities for several general classes of time-varying channels (block-memoryless, asymptotically block-memoryless, ISI, etc.) under causal CSI assumption (perfect or imperfect), as well as a list of references of the capacities for multi-input multi-output (MIMO) fading channels.

In this paper, we present a capacity-achieving communication scheme for a single-input single-output (SISO) FSMC with additive white Gaussian noise (AWGN), under the assumption of delayed noiseless output feedback and DTRCSI, subject to an *average* channel input power constraint. We consider the case without ISI. Although the access to channel outputs by the transmitter cannot improve capacity in this scenario [4], we show that it leads to simpler encoders and decoders, shortens coding delays, and has a doubly exponential decay of the probability of decoding error, while achieving any rate below the feedback capacity. This scheme is among the first to achieve the feedback capacity of a *fading* channel with AWGN.

Our scheme may be seen as a *nontrivial* extension of both the Schalkwijk-Kailath scheme (SK scheme) for the (non-fading) AWGN channel with output feedback [10], [11] and the optimal communication scheme over an FSMC without output feedback [2], [4]. In essence, our optimal feedback communication design for an FSMC consists of a set of *decoupled* subsystems running in parallel, and the subsystems are multiplexed to share the forward channel and feedback channel according to the channel state evolution. This provides the feedback communication system with the crucial ability to *adapt* its operations to the channel variation. Though the multiplexing idea was widely studied in communication without feedback, it has not received sufficient attention in the feedback communication literature. This is somewhat due to the fact that in feedback communication systems a considerably more complicated multiplexing design is often needed: While in communication without feedback, the subsystems are naturally decoupled from each other, in communication with delayed feedback, the decoupling of subsystems is possible only through appropriate designs. It turns out that, to ensure decoupling and the optimality in the feedback case, each subsystem needs to depend on an "augmented" channel state. In the case of m possible channel state values with one-step delayed output feedback, the transmitter needs to switch among m^2 possible sets of parameters, as opposed to the m required in the multiplexing design for communication without feedback ([2], [4]).

Our design simplifies when the channel states form an i.i.d. process taking a finite number of values, and requires only a simple first-order transmitter and receiver, without the needs of multiplexing or channel state augmentation. We also show that the design extends to the case when the channel states are i.i.d. taking an uncountably infinite number of values, including as special cases the widely used Rayleigh, Rician, Nakagami, and Weibull fading channel models.

Besides the novel scheme and its information theoretic properties, the other main contribution of this paper lies in the approach followed to derive the scheme, which is based on the equivalence between the feedback communication problem over an FSMC and a related feedback stabilization problem that has the FSMC in the loop. In particular, we show that the proposed communication scheme is associated with certain optimal stabilization problem of a Markov Jump Linear System (MJLS) over the FSMC. In fact, the control system framework is quite simpler to understand and design, and from it, the much less intuitive feedback communication system can be simply derived. To be more specific, if the MJLS, unstable in open loop, is stabilized in closed loop, then the communication system can transmit at a signalling rate with asymptotically vanishing probability of error (i.e., *stabilization implies reliable communication*), and the supremum signalling rate is solely determined by how fast the state of the MJLS

grows in open loop [1] . The transmission power in the communication system can also be determined from the MJLS by solving an optimal control problem called *cheap control* of the minimum variance linear quadratic Gaussian (LQG) problem. That is, the optimality in the communication problem coincides with that in the control problem.

We refer to [13] for the study of MJLS; [12], [14]–[16], and Sec. III-A and Sec. III-E in the present paper for the cheap control, or the closely related minimum-energy control, over time-invariant channels; [12], [16]–[26] for some studies on the interactions between information and control. For completeness, we present a rather brief review of the interactions between information and control. [20] formulates the feedback communication problem as a stochastic optimal control problem, and provided a dynamical programming based solution. [23] shows the fundamental connections between the communication of non-stationary, non-ergodic sources and the stabilization of unstable systems. [12] establishes, over a time-invariant Gaussian channel, the equivalence of feedback communication and feedback stabilization problems, and that the optimality in the two problems coincides. The present paper is mainly along the line of [12] and is focused on the extension to time-varying Gaussian channels.

We remark that the control-oriented feedback communication design approach, in the simplest case of AWGN channels, can give rise to an SK-type scheme. Compared with approaches attempting to directly extend the SK scheme, the control-oriented approach has been shown as a more systematic though simplified way in addressing a variety of feedback communication problems as the control systems associated with the communication schemes are usually easier to work with; see [27] for a survey. On the other hand, many feedback communication schemes have been derived by extending the SK scheme, more or less directly. For a partial list of the feedback communication designs based on the idea of Schalkwijk and Kailath, see [12], [18], [23], [27]–[35] and therein references.

This paper is organized as follows. Section II introduces the channel model and the problem we want to solve. Section III describes the proposed feedback communication scheme. This scheme is shown to achieve any rate below the capacity in Section IV. In Section V we present a numerical example. In Section VI we conclude the paper. The presentation of the above results is focused on the case of one-step-delayed feedback; the extension to the case of multi-step-delayed feedback, having a much more complex form but bearing essentially the same idea of introducing appropriate multiplexing and augmented channel states, is skipped in this paper for brevity (see [27] for details).

Notations: We represent time indices by subscripts, such as A_n; to conform with the convention in dynamical systems, the time index starts from 0. We denote by A_n^m the sequence $\{A_n, A_{n+1}, \cdots, A_m\}$, and $\{A_k\}$ the infinite sequence $\{A_k\}_{k=0}^{\infty}$. We use boldface letter $\boldsymbol{x}$ for a vector, and $x^{(i)}$ for the ith element of vector $\boldsymbol{x}$. Note that A_n^m is a sequence, $(A_n)^m$ is the mth power of A_n, $\boldsymbol{A}_n$ is a vector with time index n, and $A_n^{(m)}$ is the mth element of vector $\boldsymbol{A}_n$. We use $a[1], a[2], \cdots$ to represent a collection of fixed numbers. We denote "defined to be" as ":=".

---

[1] We will show that the *supremum achievable signalling rate* in the communication system equals the average open-loop growth rate in the MJLS. When the MJLS is open-loop unstable, the system state $\boldsymbol{x}_k$ grows as the time index k increases. The average open-loop growth rate may be defined as $\lim_{k \to \infty} (1/k) \log | \prod_{j=1}^{m} (x_k^{(j)}/x_0^{(j)}) |$, where $\boldsymbol{x}_k$ has m dimensions and $x_k^{(j)}$ denotes the jth dimension. In the case of a scalar unstable LTI system defined as $x_{k+1} = ax_k$ and $y_k = cx_k$ with $|a| > 1$, the growth rate becomes $\log |a|$, which is equal to the *degree of instability* introduced in [12]. Note that in [12] the degree of instability for a multi-dimensional time-invariant system $\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k$ is defined as $\log | \prod_i \lambda_{u,i}(A) |$, where $\lambda_{u,i}(A)$ denotes an unstable eigenvalue of A; the degree of instability is shown to equal the supremum achievable rate in the associated communication system in [12]. In this paper we will extend this notion to *multi-dimensional time-varying systems*. See Sec. III-A and Sec. III-E for more discussions.

II. CHANNEL MODEL

Fig. 1 (a) shows the forward channel considered in this paper, that is, an FSMC with AWGN, or AFSMC for short. At time k, this discrete-time channel $\mathcal{F}$ is described as

$$\mathcal{F}: \quad y_k = S_k u_k + N_k, \quad \text{for } k = 0, 1, 2, \cdots, \tag{1}$$

where u_k is the channel input, S_k is the channel state, N_k is the channel noise, and y_k is the channel output. These variables are real-valued. The noise $\{N_k\}$ is independent Gaussian with zero mean and a unit variance. The channel state S_k is independent of the channel inputs u_0^k and outputs y_0^{k-1} when conditioned on the previous states S_0^{k-1}. Furthermore, $\{S_k\}$ is a stationary, irreducible, aperiodic, finite-state homogeneous Markov chain and hence is ergodic. The one-step transition probability is

$$p_{ij} := \Pr(S_k = s[j] | S_{k-1} = s[i]), \quad \text{for } k = 1, 2, \cdots,$$

where $i, j = 1, 2, \cdots, m$; m is the number of possible state values of the Markov chain; and $s[i]$ is a fixed number for each i. Assume that $s[i] \neq s[j]$ if $i \neq j$. Note that $s[i]$ denotes one of the m states of the Markov chain, and also represents the associated channel gain if the channel is in that state. Denote the stationary distribution vector of the Markov chain (which by ergodicity exists and is unique) as $\boldsymbol{\pi} := [\pi[1], \pi[2], \cdots, \pi[m]]$.

Definition 1. *i) Define the set of all possible channel state sequences $\{S_k\}$ as Ω, i.e.,*

$$\Omega := \{\{S_k\}\}. \tag{2}$$

ii) Define the set of all possible channel state sequences S_0^k as Ω_k, i.e.,

$$\Omega_k := \left\{ S_0^k \right\}. \tag{3}$$

iii) Define the (k, μ)-typical set of sequences S_0^k as

$$\Omega_{k,\mu} := \left\{ S_0^k \left| \left| \frac{n(j, k)}{k+1} - \pi[j] \right| < \mu \text{ and } \left| \frac{n(j, l, k)}{n(j, k)} - p_{jl} \right| < \mu \right. \right\}, \tag{4}$$

where $j, l = 1, 2, \cdots, m$; $\mu > 0$; and for a given channel state sequence S_0^k, $n(j, l, k)$ is the number of transitions from channel state $s[j]$ to channel state $s[l]$ up to time k, and $n(j, k) := \sum_{l=1}^{m} n(j, l, k)$. [2]

Note that the typicality we defined above is in fact strong typicality [36]. By the Weak Law of Large Numbers or (strong) Asymptotic Equipartition Property (AEP), it holds that

$$\Pr(\Omega_{k,\mu}) \to 1 \tag{5}$$

as k tends to infinity, and that

$$
\begin{aligned}
\frac{n(j, k)}{k+1} &\xrightarrow{\text{P}} \pi[j] \\
\frac{n(j, l, k)}{n(j, k)} &\xrightarrow{\text{P}} p_{jl} \\
\frac{n(j, l, k)}{k+1} &\xrightarrow{\text{P}} \pi[j] p_{jl}.
\end{aligned}
\tag{6}
$$

Here $\xrightarrow{\text{P}}$ specifies convergence in probability. Note that if $a_k \xrightarrow{\text{P}} a$ and $b_k \xrightarrow{\text{P}} b$, then $a_k b_k \xrightarrow{\text{P}} ab$ and $f(a_k) \xrightarrow{\text{P}} f(a)$ for any continuous function f; see Lemma 3.3 (i.e. the Continuous Mapping Theorem) and Corollary 3.5 in [37].

---

[2] To simplify notations, we do not specify the dependency of $n(j, l, k)$ and $n(j, k)$ on the given sequence S_0^k here. If S_0^k is viewed as a realization, then $n(j, l, k)$ and $n(j, k)$ are not random variables, in which case one may use the notations $S_0^k(\omega_0)$, $n(j, l, k)(\omega_0)$, and $n(j, k)(\omega_0)$ where ω_0 is one fixed sample point in the sample space Ω_k. If, however, S_0^k is viewed as a random vector, then $n(j, l, k)$ and $n(j, k)$ are to be viewed as random variables, in which case one may use the notations $S_0^k(\omega)$, $n(j, l, k)(\omega)$, and $n(j, k)(\omega)$ that are functions of ω defined on the sample space Ω_k.

We mainly focus on channel $\boldsymbol{F}$ with the following assumptions in this paper:

Definition 2. *Define an AFSMC $\mathcal{F}$ with one-step-delayed output feedback and DTRCSI as channel $\boldsymbol{F}$.*

In other words, we consider the case of one-step-delayed transmitter-side and instantaneous receiver-side CSI (DTRCSI). We also allow the transmitter to have access to the one-step-delayed channel output, i.e., the receiver at time k having observed y_0^k will compute v_k (depending only on y_0^k) and feed back v_k along with S_k to the transmitter with one step delay. In other words, the channel input u_k can depend on S_0^{k-1} and v_0^{k-1}. See Fig. 1 (b).
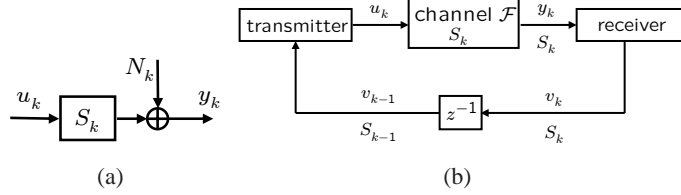

Fig. 1.   (a) An AFSMC $\mathcal{F}$. (b) A system over channel $\boldsymbol{F}$.

The capacity for channel $\boldsymbol{F}$ is characterized in [4] as

$$C = \max_{\Pr(u_k|S_{k-1})} \mathbf{E}_{S_{k-1}\sim\boldsymbol{\pi},S_k} I(u_k; y_k|S_{k-1}, S_k), \tag{7}$$

where $\Pr(u_k|S_{k-1})$ is any input distribution subject to the *average* channel input power constraint

$$\mathbf{E}u_k^2 \leq \mathcal{P}. \tag{8}$$

Note that S_k also follows the stationary distribution $\boldsymbol{\pi}$ since S_{k-1} does. This optimization problem can be reduced to

$$C = \max_{\Gamma(\cdot):\mathbf{E}_{S_{k-1}}\Gamma(S_{k-1})\leq\mathcal{P}} \frac{1}{2}\mathbf{E}_{S_{k-1}\sim\boldsymbol{\pi},S_k} \log(1 + (S_k)^2\Gamma(S_{k-1})), \tag{9}$$

where $\Gamma(\cdot)$ is the power allocation function that maps the channel state S_{k-1} to the channel input power $\Gamma(S_{k-1})$. The above capacity formula is obtained by invoking Lemma 2 of [4] (with $d = 1$ and $\sigma_s^2 = 1$ therein). The *optimal* power allocation, denoted $\gamma(\cdot)$, is given by the solution of a set of m equations [3] derived by applying the Kuhn-Tucker condition (see Appendix B in [4]) and is assumed given throughout this paper. The objective of the paper is to design a transmitter and receiver for channel $\boldsymbol{F}$ to achieve any rate below the capacity given in (9).

Note that the above defined channel has a discrete channel state but continuous channel input, noise, and output (in contrast to the discreteness of FSMC inputs and outputs assumed in [1]–[6], with the notable exception of some parts in [4]). This channel may be used in modelling the following cases and their generalizations. For one, a channel is subject to both erasure (i.e. discrete channel states) and AWGN (i.e. continuous noise), and the erasures may exhibit certain time correlation (e.g. forming a 2-state Markov chain) as the causes of erasures may be time-correlated in some cases (such as buffer overflow). For another, a channel is subject to bursty noises with different noise variances, and the occurrence of bursty noises forms a finite-state Markov chain. The well-known Gilbert-Elliot channel with AWGN falls into this category.

---

[3]The equations are not linear but the nonlinearity involves fractions only. Hence, they can be easily solved numerically. Besides, the decision variables $\Gamma(s[i])$, $i = 1, \cdots, m$, are inside a compact region and a number of numerical approaches, such as branching-and-bound, are available to improve the search efficiency.

We also remark that the assumption of the instantaneous, prefect CSI at the receiver side, though often assumed in the literature (see [2], [4], [38], etc.), is not quite realistic (especially in the fast fading case); we adopt this assumption in order to simplify the analysis and to gain some conceptual understandings of the feedback communication problem over Markov channels. A study taking into consideration of the imperfect CSI will be subject to future work, and our present research based on perfect CSI may be found useful in that study.

III. THE PROPOSED FEEDBACK COMMUNICATION SCHEME

In this section, we propose a communication scheme for channel $\boldsymbol{F}$. After a brief review of a feedback communication design for an AWGN channel and a description of the main idea of our design, we introduce the communication scheme and some of its properties. This scheme will be shown to achieve any rate below the feedback capacity in Section IV.

A. Review of an SK-type communication system for an AWGN channel with feedback

To better explain the proposed scheme, we review a feedback communication scheme over an AWGN channel, see e.g. [10]–[12], [18], [39] for more details. Fig. 2 shows the communication system which can achieve any rate below the Shannon capacity of the AWGN channel. The coding process is as follows. Fix a coding length $(K+1)$, a power budget $\mathcal{P} > 0$, and any $\epsilon > 0$ (where ϵ is an arbitrarily small slack from the Shannon capacity). Define

$$a := \sqrt{1+\mathcal{P}}, \quad b := a - \frac{1}{a}, \quad c := 1, \tag{10}$$

and

$$M_K := a^{(K+1)(1-\epsilon)}. \tag{11}$$

Equally partition the interval $[-\frac{1}{2}, \frac{1}{2}]$ into M_K sub-intervals, and map the sub-interval centers to a set of M_K equally likely messages; this is known to both the transmitter and receiver *a priori*. Suppose that we wish to transmit the message represented by the center W. Let $y_{-1} := 0$ and $x_{-1} := W/a$, i.e., $x_0 := W$. In other words, *the initial condition (at time 0) of the transmitter is the to-be-transmitted message*. Generate $\hat{x}_{0,k}$ according to the following dynamics

$$
\begin{aligned}
x_k &= a x_{k-1} - b y_{k-1} \\
u_k &= c x_k \\
y_k &= u_k + N_k \\
\hat{x}_{0,k} &= \hat{x}_{0,k-1} + a^{-k-1} b y_k.
\end{aligned}
\tag{12}
$$

We then decode, at the decoder side, by mapping $\hat{x}_{0,K}$ into the closest sub-interval center.

One can show that the probability of decoding error vanishes as k tends to infinity. In addition, the average input power $\mathbf{E}(u_0^{k\prime} u_0^k)/(k+1)$ converges to $\mathcal{P}$. Since

$$R := \lim_{K \to \infty} \frac{1}{K+1} \log M_K = (1 - \epsilon) \log a = \frac{1 - \epsilon}{2} \log(1 + \mathcal{P}), \tag{13}$$

it holds that any rate below the Shannon capacity is achievable.

We can also transmit a Gaussian random variable over the AWGN channel. Suppose $x_0 \sim \mathcal{N}(0, \mathcal{P})$ (i.e. $x_{-1} \sim \mathcal{N}(0, \mathcal{P}/a^2)$). Following the choice of parameters in (10) and the dynamics in (12), one obtains that

$$\mathrm{MSE}(\hat{x}_{0,k-1}) := \mathbf{E}(x_0 - \hat{x}_{0,k-1})^2 = \mathcal{P} a^{-2k}. \tag{14}$$

This is indeed the *minimum* mean-squared error (MSE) that we can achieve over this channel, since this channel can transmit at a rate no higher than $\log a$, which corresponds to the (minimum) MSE distortion $\mathcal{P} a^{-2k}$. In other words, x_0 is *successively refined* at the receiver at the Shannon capacity rate; see e.g. [18], [23], [40] for related study.
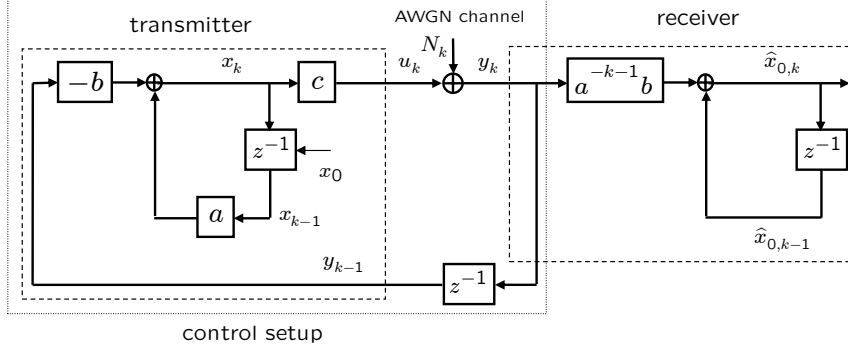
6

Fig. 2. The communication system for an AWGN channel. $x_{-1} = W/a$, $x_0 = W$, $\hat{x}_{0,0} = 0$, $N_k \sim \mathcal{N}(0,1)$, $a > 1$, and $k \geq 0$. The system inside the dotted box represents a closed-loop control system.

The optimality in both the digital and analog communication problems is associated with the optimality in a control problem over the control setup as indicated in Fig. 2. Suppose in the control setup, for any fixed $a > 1$ and c, we wish to minimize the power of u by appropriately designing b (noticing that when the power of u is finite then the control setup is stabilized in closed loop) according to the solution to a classical control problem known as *cheap control* of minimum variance LQG problem ([12], [14], [41], [42]). Precisely, the dynamics of the scalar LTI control system is

$$
\begin{aligned}
x_k &= ax_{k-1} - by_{k-1} \\
u_k &= cx_k \\
y_k &= u_k + N_k
\end{aligned}
\tag{15}
$$

where x_k is the system state with an unknown initial condition x_0, u_k is the system output, y_k is the measurement of the system output (corrupted by noise N_k) and the input to the controller, $(-b)$ is the controller gain, and $(-by_k)$ is the controller's output (also known as the control input to the system that drives the system state). Here a and c are given, and we need to find $(-b)$ to minimize the output power. That is, we need to solve the following control problem

$$
\min_{b} \quad \lim_{k \to \infty} \frac{1}{k+1} \mathbf{E} \sum_{t=0}^{k} (u_t)^2,
\tag{16}
$$

in which the control effort is free as there is no penalty on the controller's output $(-by_k)$; hence the name "cheap control" [4].

Note that $x_k = a^k x_0$ when the control system is in open loop (i.e. $b = 0$), and hence, we may define the growth rate for this unstable system as

$$
\frac{1}{k} \log \frac{x_k}{x_0} = \log a.
\tag{17}
$$

[4] This cheap control problem may be reformulated as an *expensive control* problem (depending on whether one treats b or c as the controller). Treating u as the controller's output, c as the state-feedback controller gain to be designed, and b as given, one obtains an equivalent optimal control problem to minimize the power of the control effort (subject to closed-loop stability); hence the name "expensive control". The noiseless version of the expensive control is a minimum-energy control problem as the control energy in infinite horizon can be made finite. In both cheap control and expensive control (as well as the minimum-energy control), the optimal controllers place the closed-loop eigenvalues at the reciprocals of open-loop eigenvalues. In [12], the expensive control formulation (as opposed to cheap control) was under study when investigating the relationship between feedback control and feedback communication.

The optimal solution to the above control problem is to choose b as $(a - 1/a)$, which places the closed-loop eigenvalue at the reciprocal of the open-loop eigenvalue [5]; the supremum communication rate is $\log a$ (cf. (13)), equal to the growth rate of the unstable open-loop of the control setup, and the cheap control guarantees that the minimum transmission power is used to achieve any rate arbitrarily close to $\log a$. These ideas were extended to feedback Gaussian channels with memory in [12], and will be extended to an AFSMC with feedback as we will show in this paper.

We point out that our formulation of the communication scheme as described above differs from other popular SK-type feedback communication schemes in their original forms. We comment on the relationship and differences between these formulations; these comments also apply to the proposed scheme for an AFSMC and its variations (see Table I). First, the above-mentioned formulation does not involve unbounded coding parameters or unbounded signal power, whereas that in [10] involves exponentially growing bandwidth, [11] involves an exponentially growing parameter α^k where $\alpha > 1$ and k denotes the time index, and [12] generates a feedback signal with exponentially growing power, despite the facts that they all generate the same channel inputs, same outputs, and same decoded messages, and that one formulation can be obtained as a simple reformulation of others. Thus we consider our formulation more feasible (at least for simulation purposes, if not practically implementable). We also remark that our formulation is essentially the scheme studied by Gallager (p. 480, [39], which may need to receive more attention in the literature) with minor differences. In addition, our formulation differs from the original SK scheme in that, ours performs the same operation at every step, whereas the original SK formulation performs its startup operation different from later steps. Although ours has the advantage of unifying the operations for all steps (which simplifies the control-oriented analysis, also cf. Section IV in [12]), it has to remove a bias term using an extra equation when used to transmit digit messages (as done in [39]; see the equation mapping v_N to r_N on p. 481) or wait long enough until that exponentially vanishing bias becomes negligible (see comments in Section IV of [12]). In contrast, the original SK scheme is unbiased since the special startup operation eliminates the bias. In this paper, we focus on our formulation and its extension since it corresponds to a control system that is easier to analyze.

B. General description of the proposed scheme

We present an informal overview of the proposed scheme before going into the technical details. In short, the proposed design can be viewed as a process of multiplexing a set of subsystems with feedback, each of them using an augmented channel state. In the degenerated case where the channel is time-invariant, the design can be simplified to the one described in the previous subsection for AWGN channels.

Suppose that the Markov channel $\mathcal{F}$ has m possible state values. Then our scheme consists of m subsystems (each of which is associated with one channel state value) sharing the forward channel $\mathcal{F}$. Represent each to-be-transmitted message as an m-dimensional codeword which contains m sub-codewords, and the m sub-codewords uniquely determine the message. Then assign the ith sub-codeword to be the initial condition of the ith subsystem (see Section III-H and Fig. 6). This completes the "encoding" process.

---

[5] We provide a sketched proof here. When (15) is stabilized in closed loop, the power $\mathbf{E}(u_k)^2$ convergences to $b^2/(1-(a-b)^2)$. Then it is straightforward to compute that $b = (a - 1/a)$ is the minimizer and the minimum power is $(a^2 - 1)$.

Then the system operates according to its designed dynamics (see (18) and Fig. 3), in a way that the m subsystems are multiplexed and each of the m sub-codewords is communicated without the interference from other sub-codewords. To ensure this, at each time epoch, one and only one subsystem is selected to use the forward channel to send its sub-codeword, based on two immediate previous channel states (i.e. an augmented channel state; see the dependence on S_{k-2} and S_{k-1} at time k in Fig. 3). The introduction of the augmented channel state leads to the decoupling of the m subsystems, which eases decoding without loss of optimality. Note that the decoupling is impossible (except for the degenerated case that $\{S_k\}$ is i.i.d.) if the transmitter only uses the immediate past channel state in the multiplexing. Note also that state augmentation is widely used for control systems with delay. See Section III-F for the details regarding the decoupling of subsystems.

To decode, each subsystem performs its decoding process independently. Then the original message would be correctly recovered if *all* m sub-codewords are correctly recovered. Therefore, the average rate of our scheme is the weighted sum of the individual rates for the subsystems, and the weights are the probabilities of the subsystems being selected to use the forward channel. Besides, the decoupling and the cheap control design (see Section III-E) ensure that the channel input power at time k depends on the channel state at time $(k-1)$ only, and the power of the ith subsystem is designed to converge to $\gamma(s[i])$, resulting in the average power converging to the weighted sum of $\gamma(s[i])$, which is the optimal power (see Sec. IV-B). To summarize, each subsystem achieves a rate arbitrarily close to its capacity by applying the corresponding cheap control solution, and the overall scheme achieves a rate arbitrarily close to the capacity of channel $\boldsymbol{F}$.

C. Communication scheme

Fig. 3 shows the proposed communication scheme. Its seemingly complicated and non-intuitive operations may be obtained easily as a simple reformulation of a control system; see Section III-E for details. In this figure, we identify the transmitter, the channel $\mathcal{F}$, and the receiver in the dashed boxes. We call $\boldsymbol{x}_k \in \mathbb{R}^m$ the *transmitter state*, with $\boldsymbol{x}_0$ being the initial condition. We call $\hat{\boldsymbol{x}}_{0,k} \in \mathbb{R}^m$ the *receiver estimate*, which is an estimate of $\boldsymbol{x}_0$ at time k. Parameters $A \in \mathbb{R}^{m \times m}$, $\boldsymbol{b} \in \mathbb{R}^m$, and $\boldsymbol{c} \in \mathbb{R}^m$ depend causally on the channel states, and will be chosen to reflect the adaptation of the communication strategy to the channel variation.
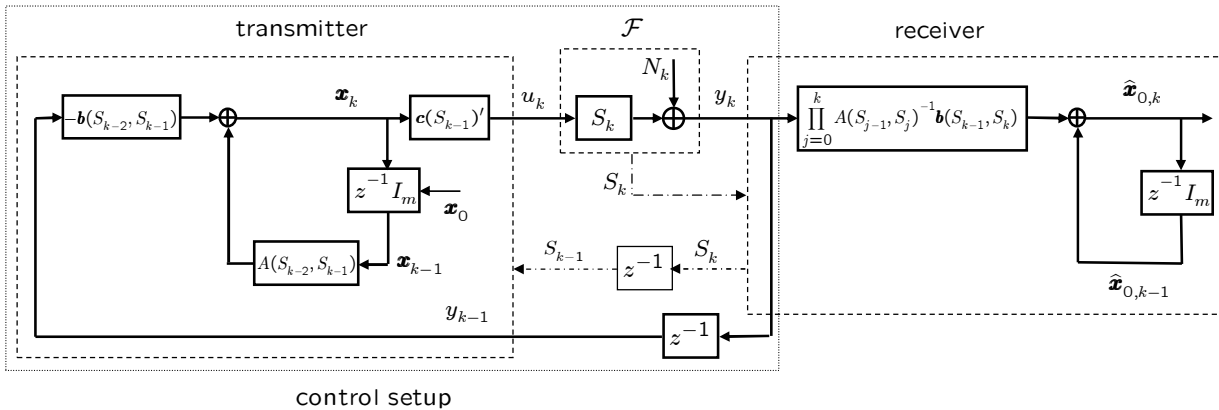

Fig. 3. The communication scheme and the control setup.

At time k, $k \geq 0$, the system generates signals according to the following dynamics in the listed

order:

$$
\begin{aligned}
\boldsymbol{x}_k &= A(S_{k-2}, S_{k-1})\boldsymbol{x}_{k-1} - \boldsymbol{b}(S_{k-2}, S_{k-1})y_{k-1} \\
u_k &= \boldsymbol{c}(S_{k-1})'\boldsymbol{x}_k \\
y_k &= S_k u_k + N_k \\
\hat{\boldsymbol{x}}_{0,k} &= \hat{\boldsymbol{x}}_{0,k-1} + \left(\prod_{j=0}^{k} A(S_{j-1}, S_j)^{-1} \right) \boldsymbol{b}(S_{k-1}, S_k)y_k,
\end{aligned}
\tag{18}
$$

where $S_{-2} := s[1]$, $S_{-1} := s[1]$, $y_{-1} := 0$, $\boldsymbol{x}_{-1} := A(S_{-2}, S_{-1})^{-1}\boldsymbol{x}_0$, and $\hat{\boldsymbol{x}}_{0,-1} := 0$. The above recursions will generate a sequence of receiver estimates $\{\hat{\boldsymbol{x}}_{0,k}\}$ that converges to $\boldsymbol{x}_0$, as we will prove in the next section.

D. Choice of parameters

Given any $\mathcal{P} > 0$, we choose the parameters in the communication setup as follows. Let $\gamma(\cdot)$ be the optimal power allocation function computed from [4]. Supposing that $S_{k-2} = s[j]$ for some j and $S_{k-1} = s[l]$ for some l, we define

$$
\begin{aligned}
A(S_{k-2}, S_{k-1}) &:= \mathrm{diag}([1, \cdots, 1, a(S_{k-2}, S_{k-1}), 1, \cdots, 1]) &&\in \mathbb{R}^{m \times m} \\
\boldsymbol{b}(S_{k-2}, S_{k-1}) &:= [0, \cdots, 0, b(S_{k-2}, S_{k-1}), 0, \cdots, 0]' &&\in \mathbb{R}^m \\
\boldsymbol{c}(S_{k-1}) &:= [0, \cdots, 0, c(S_{k-1}), 0, \cdots, 0]' &&\in \mathbb{R}^m,
\end{aligned}
\tag{19}
$$

where $a(S_{k-2}, S_{k-1})$ is the (j, j)th element of $A(S_{k-2}, S_{k-1})$, given by

$$
a(S_{k-2}, S_{k-1}) := \sqrt{(S_{k-1})^2 \gamma(S_{k-2}) + 1} \quad ;
\tag{20}
$$

$b(S_{k-2}, S_{k-1})$ is the jth element of $\boldsymbol{b}(S_{k-2}, S_{k-1})$, given by

$$
\begin{aligned}
b(S_{k-2}, S_{k-1}) &:= \frac{\gamma(S_{k-2})S_{k-1}}{\sqrt{(S_{k-1})^2 \gamma(S_{k-2}) + 1}} = \frac{\gamma(S_{k-2})S_{k-1}}{a(S_{k-2}, S_{k-1})} \\
&= \frac{1}{S_{k-1}} \left(a(S_{k-2}, S_{k-1}) - \frac{1}{a(S_{k-2}, S_{k-1})} \right),
\end{aligned}
\tag{21}
$$

where the last equality is also true for the case $S_{k-1} = 0$ if we treat $0/0 = 0$; and $c(S_{k-1})$ is the lth element of $\boldsymbol{c}(S_{k-1})$, given by

$$
c(S_{k-1}) := 1.
\tag{22}
$$

Whenever S_k, $k < 0$, is encountered, it is treated as $s[1]$. Note that the above choice of A and $\boldsymbol{b}$ uses the augmented channel state (S_{k-2}, S_{k-1}) as we have mentioned in Section III-B.

We can rewrite the dynamics of the transmitter state $\boldsymbol{x}_k$ in (18) as

$$
\boldsymbol{x}_k = A_{cl}(S_{k-2}, S_{k-1})\boldsymbol{x}_{k-1} - \boldsymbol{b}(S_{k-2}, S_{k-1})N_{k-1},
\tag{23}
$$

where

$$
A_{cl}(S_{k-2}, S_{k-1}) := A(S_{k-2}, S_{k-1}) - S_{k-1}\boldsymbol{b}(S_{k-2}, S_{k-1})\boldsymbol{c}(S_{k-2})'
\tag{24}
$$

is the closed-loop matrix for generating $\boldsymbol{x}_k$. Let us assume $S_{k-2} = s[j]$. With the above choice of parameters, we then obtain that $A_{cl}(S_{k-2}, S_{k-1})$ in (24) is a *diagonal* matrix whose (i, i) element is 1 if $i \neq j$, and is

$$
a(S_{k-2}, S_{k-1}) - S_{k-1}b(S_{k-2}, S_{k-1})c(S_{k-2}) = a(S_{k-2}, S_{k-1})^{-1}
\tag{25}
$$

if $i = j$. Hence, we have

$$
x_k^{(i)} = \begin{cases} a(S_{k-2}, S_{k-1})^{-1}x_{k-1}^{(i)} - b(S_{k-2}, S_{k-1})N_{k-1} & \text{if } i = j \\ x_{k-1}^{(i)} & \text{if } i \neq j; \end{cases}
\tag{26}
$$

10

or equivalently in matrix form (noticing that $A_{cl}(S_{k-2}, S_{k-1}) = A(S_{k-2}, S_{k-1})^{-1}$ [6])

$$\boldsymbol{x}_k = A(S_{k-2}, S_{k-1})^{-1}\boldsymbol{x}_{k-1} - \boldsymbol{b}(S_{k-2}, S_{k-1})N_{k-1}. \tag{27}$$

More explicitly,

$$
\begin{bmatrix} x_k^{(1)} \\ \vdots \\ x_k^{(j)} \\ \vdots \\ x_k^{(m)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & & \cdots & & 0 & 0 \\ 0 & 1 & 0 & & \cdots & & 0 & 0 \\ \cdots & \cdots & & & & & & \\ 0 & 0 & \cdots & a(S_{k-2}, S_{k-1})^{-1} & \cdots & 0 & & \\ \cdots & \cdots & & & & & & \\ 0 & \cdots & & 0 & & 1 & 0 \\ 0 & \cdots & & 0 & & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{k-1}^{(1)} \\ \vdots \\ x_{k-1}^{(j)} \\ \vdots \\ x_{k-1}^{(m)} \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b(S_{k-2}, S_{k-1}) \\ 0 \\ \vdots \\ 0 \end{bmatrix} N_{k-1} \tag{28}
$$

This illustrates that, conditioned on the channel state sequence S_0^k, the evolution of the subsystem of $x_k^{(i)}$ does not involve $x_0^{(l)}$ for any $l \neq i$.

For future references, we define

$$\bar{a}[j] := \prod_{l=1}^{m} a(s[j], s[l])^{\pi[j]p_{jl}}, \quad \text{for } j = 1, \cdots, m \tag{29}$$

and

$$\tilde{a} := \prod_{j=1}^{m} \bar{a}[j]. \tag{30}$$

E. Cheap control

Equations (23) and (24) specify a *control system*, referred to as the *control setup*, in which we want to minimize the average power of u, namely the channel input power, by designing $\boldsymbol{b}$ for a given A and $\boldsymbol{c}$. This is a *cheap control problem over a Markov channel*. The presence of cheap control in the optimal communication scheme is necessary because for any fixed, unstable matrix $A(S_{k-2}, S_{k-1})$, the supremum signalling rate is fixed under certain stability conditions. This follows that one needs to minimize the average channel input power by design, namely the average power of $u_k = \boldsymbol{c}(S_{k-1})'\boldsymbol{x}_k$, a cost quadratic in $\boldsymbol{x}_k$, which is easily reformulated as a cheap control problem. The resulting (minimum) average power of u_k will be computed in Proposition 2. In this subsection, we briefly discuss the cheap control problem.

In Fig. 4 we illustrate how we obtain the optimal communication scheme of Fig. 3 from the cheap control. By linearity, it holds that

$$\boldsymbol{x}_k = \tilde{\boldsymbol{x}}_k + \hat{\boldsymbol{x}}_k, \tag{31}$$

where $\tilde{\boldsymbol{x}}_k$ is the zero-input response (generated purely from initial condition $\boldsymbol{x}_0$), and $\hat{\boldsymbol{x}}_k$ is the zero-state response (generated purely from external input y_0^k); this is shown in the left part of Fig. 4 (b). Note that $-\hat{\boldsymbol{x}}_k$ can be alternatively constructed from y_0^k as indicated in the right part of Fig. 4 (b). Since $\boldsymbol{x}_k$ is bounded due to the closed-loop stability and since $\tilde{\boldsymbol{x}}_k$ grows "exponentially" (in a time-varying manner), it holds that

$$\tilde{\boldsymbol{x}}_k \approx -\hat{\boldsymbol{x}}_k \tag{32}$$

---

[6]One can see that, when a subsystem is activated, the closed-loop eigenvalue of this subsystem is placed at the reciprocal of the open-loop eigenvalue, which resembles the cheap control design for a Gaussian channel without fading [14].

for sufficiently large k. Therefore, without actually knowing $\boldsymbol{x}_0$ beforehand, the right part of Fig. 4 (b) can use $-\hat{\boldsymbol{x}}_k$ to approximate $\tilde{\boldsymbol{x}}_k$ and hence reconstruct $\boldsymbol{x}_0$ using $\hat{\boldsymbol{x}}_{0,k} := \prod_{j=0}^{k} A(S_{j-1}, S_j)^{-1}(-\hat{\boldsymbol{x}}_k)$, provided that the knowledge of $\{S_k\}$ is known [7]. Then block diagram transformations (that shifts the one-step delay into the feedback link) lead to the proposed communication scheme in Fig. 3 which can be used to convey $\boldsymbol{x}_0$. This is how the seemingly complicated communication scheme is obtained.
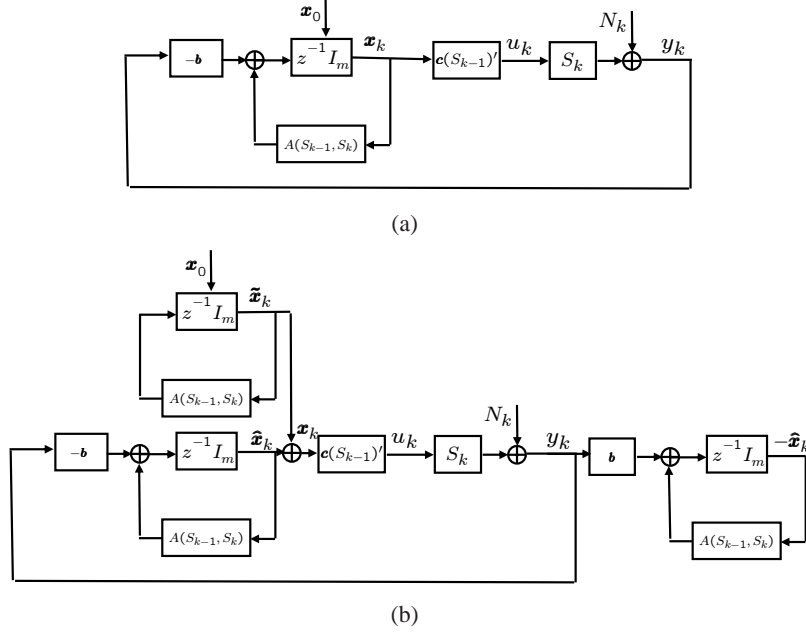

(a)

(b)

Fig. 4. (a) Cheap control of the MJLS. (b) Intermediate step towards the optimal communication scheme.

Below we discuss some additional properties of the cheap control problem. Consider the cheap control of the MJLS shown in Fig. 4 (a). More specifically, the closed-loop control system is

$$
\begin{aligned}
\boldsymbol{x}_{k+1} &= A(S_{k-1}, S_k)\boldsymbol{x}_k - \boldsymbol{b}y_k \\
u_k &= \boldsymbol{c}(S_{k-1})'\boldsymbol{x}_k \\
y_k &= S_k u_k + N_k,
\end{aligned}
\tag{33}
$$

in which $A(S_{k-1}, S_k)$ and $\boldsymbol{c}(S_{k-1})$ are given for any k; y_k is the noisy measurement of the system output and the input to the controller; $\boldsymbol{b}$ is the controller gain to be designed to ensure the minimum average power of u, with the discrete state S_0^k known to the controller at time k; and $\boldsymbol{b}y_k$ is the controller's output (also known as the control input to the system that drives the system state). Namely, we consider the MJLS with noisy observation y_{k-1} and perfect knowledge of the Markov state S_0^k, and the associated optimal control problem is

$$
\min_{\boldsymbol{b}} \quad \lim_{k \to \infty} \frac{1}{k+1} \mathbf{E} \sum_{t=0}^{k} (u_t)^2.
\tag{34}
$$

[7]An alternative way to see the convergence of $\hat{\boldsymbol{x}}_{0,k}$ to $\boldsymbol{x}_0$ is to combine Lemma 1 with, first, the stability of the system for $\boldsymbol{x}_k$, and second, the fact that Φ_k is vanishing (Lemma 2 i) and ii)).

Recall that $A(S_{k-2}, S_{k-1})$ is a diagonal matrix, so if all of it diagonal elements, in absolute values, are no smaller than 1 and some of the diagonal elements are strictly greater than 1, then (33) is open-loop unstable. Similar to the AWGN unstable case, for any given initial condition $\boldsymbol{x}_0$, we may define the average growth rate of $\boldsymbol{x}_k$ in open loop (i.e., $\boldsymbol{b} = 0$ and hence $\boldsymbol{x}_k = A(S_{k-2}, S_{k-1})\boldsymbol{x}_{k-1}$) as

$$
\begin{aligned}
&\lim_{k \to \infty} \frac{1}{k} \log \frac{|\prod_{j=1}^{m} x_k^{(j)}|}{|\prod_{j=1}^{m} x_0^{(j)}|} \\
&= \lim_{k \to \infty} \frac{1}{k} \log |\prod_{j=1}^{m} (\phi_k^{(j)})^{-1}| \\
&= \lim_{k \to \infty} \frac{1}{k} \sum_{j=1}^{m} \sum_{l=1}^{m} n(j, l, k) \log a(s[j], s[l]) \qquad (35) \\
&\overset{(a)}{=} \sum_{j=1}^{m} \sum_{l=1}^{m} \pi[j] p_{jl} \log a(s[j], s[l]) \\
&\overset{(b)}{=} \mathbf{E}_{S_{k-2} \sim \boldsymbol{\pi}, S_{k-1}} \log \det(A(S_{k-2}, S_{k-1})) \\
&= \log \tilde{a},
\end{aligned}
$$

where $\tilde{a}$ is defined in (30), (a) is to be interpreted as convergence in probability, (b) follows from the definition of $A(S_{k-2}, S_{k-1})$, and the last equality is shown in Appendix A. In other words, the average growth rate of the state is equal to the expected degree of instability of $A(S_{k-2}, S_{k-1})$.

In the special case that S_k is constant throughout, (34) and (35) reduce to (16) and (17), the counterparts for cheap control over an AWGN channel. If A and $\boldsymbol{c}$ are given according to (19), (20), and (22), then $\boldsymbol{b}$ according to (19) and (21) is the *optimal* choice. Instead, if the choice of A and $\boldsymbol{c}$ is *not* given but a constraint of A is present, i.e., the average growth rate of the open-loop system is given, then it can be shown that A and $\boldsymbol{c}$ given in (19), (20), and (22) will lead to the minimum power, subject to the above constraint. Detailed computation is omitted here, but the validity will become clear when combining 1) the relation between the control problem and the communication problem shown below and 2) the optimality in the communication problem established in the next section.

F. Decoupling of states

To capture the decoupling property presented in Sec. III-D precisely, we provide a lemma after introducing some notations.

Fix a channel state sequence S_0^k. If $S_{\tau-1} = s[j]$ and $S_\tau = s[l]$, we have

$$
A(S_{\tau-1}, S_\tau)^{-1} = \mathrm{diag}\left([1, \cdots, 1, a(s[j], s[l])^{-1}, 1, \cdots, 1]\right), \qquad (36)
$$

and we can then obtain

$$
\begin{aligned}
\prod_{\tau=0}^{k} A(S_{\tau-1}, S_\tau)^{-1} &= \mathrm{diag}\left(\left[\prod_{l=1}^{m} a(s[1], s[l])^{-n(1,l,k)}, \cdots, \prod_{l=1}^{m} a(s[m], s[l])^{-n(m,l,k)}\right]\right) \\
&:= \mathrm{diag}\left(\left[\phi_k^{(1)}, \cdots, \phi_k^{(m)}\right]\right) := \Phi_k,
\end{aligned}
\qquad (37)
$$

where $\phi_k^{(j)}$ and Φ_k are defined in the obvious manner; recall Definition 1 for notation $n(i, l, k)$.

Lemma 1. *Fix any channel state sequence S_0^k and initial condition $\boldsymbol{x}_0$. Then*

$$\begin{aligned}
\boldsymbol{x}_{k+1} &= \Phi_k \boldsymbol{x}_0 - \Phi_k \sum_{\tau=0}^{k} (\Phi_\tau)^{-1} \boldsymbol{b}(S_{\tau-1,\tau}) N_\tau \\
\hat{\boldsymbol{x}}_{0,k} &= \boldsymbol{x}_0 - \Phi_k \boldsymbol{x}_{k+1} \\
&= (I - (\Phi_k)^2)\boldsymbol{x}_0 + (\Phi_k)^2 \sum_{\tau=0}^{k} (\Phi_\tau)^{-1} \boldsymbol{b}(S_{\tau-1,\tau}) N_\tau \\
\hat{x}_{0,k}^{(j)} &= (1 - (\phi_k^{(j)})^2) x_0^{(j)} + (\phi_k^{(j)})^2 \left[\sum_{\tau=0}^{k} (\Phi_\tau)^{-1} \boldsymbol{b}(S_{\tau-1,\tau}) N_\tau \right]^{(j)},
\end{aligned}$$

(38)

where in the last equation, $[\cdot]^{(j)}$ extracts the jth entry of the vector.

Proof: The first equation is proved by recursively applying (27) and (37). For the second equation, straightforward computation shows that

$$\begin{aligned}
\boldsymbol{x}_{k+1} &= (\Phi_k)^{-1} \boldsymbol{x}_0 - (\Phi_k)^{-1} \sum_{\tau=0}^{k} \Phi_\tau \boldsymbol{b}(S_{\tau-1,\tau}) y_\tau \\
\hat{\boldsymbol{x}}_{0,k} &= \sum_{\tau=0}^{k} \Phi_\tau \boldsymbol{b}(S_{\tau-1,\tau}) y_\tau.
\end{aligned}$$

(39)

Finally, the third equation follows from the fact that $(I - \Phi_k)^2$ is a diagonal matrix. $\square$

From the last equation, we see that each sub-codeword $x_0^{(j)}$ is transmitted independently from other sub-codewords. To help the reader better understand the decoupling of subsystems, let us call the subsystem associated with $x_k^{(j)}$ as the *jth subsystem*, denoted $\Sigma_{s[j]}$. If it holds that $S_k = s[j]$, then we can also use the notation Σ_{S_k} (i.e. $\Sigma_{S_k} = \Sigma_{s[j]}$ if $S_k = s[j]$). The jth subsystem typically goes through the following cycle of operations: holding — updating receiver estimate (if the immediate past channel state was $s[j]$) — updating transmitter state (if the second immediate past channel state was $s[j]$) — holding. Consequently, any two subsystems that are not on hold do not perform the same updating operation at the same time. That is, our design ensures mutually exclusive updates among subsystems and hence the interaction-free evolution for subsystems. This simplifies the encoding/decoding processes.

We point out that the decoupling is not possible if an augmented channel state is not used. See Fig. 5 for the schematic of the multiplexing scheme. At time k, if, for some $t := t(k)$, the transmitter selects Σ_{S_t} to use the forward link, then the receiver should select Σ_{S_t} to receive to avoid interference. If, however, for some $\tau := \tau(k)$, the receiver selects Σ_{S_τ} to use the feedback link, then the transmitter should select $\Sigma_{S_{\tau-1}}$ to receive to avoid interference. No matter how one may pick t and τ, at least two channel states are needed.
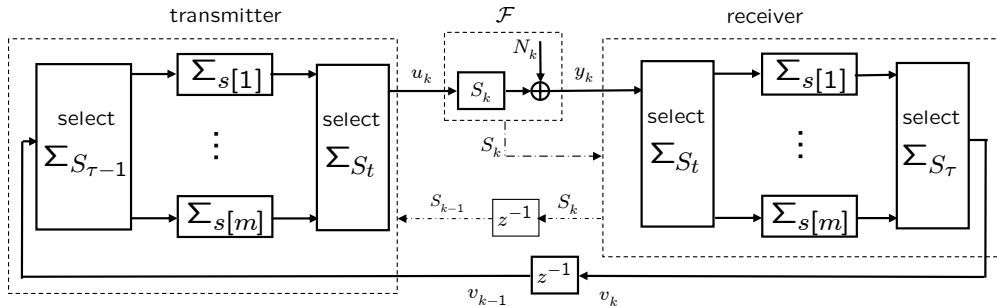

Fig. 5. The schematic of the multiplexing scheme.

The proposed scheme is in fact such that $t := \tau := k - 1$. Therefore, at time k, the transmitter sends information for $\Sigma_{S_{k-1}}$, and receives for $\Sigma_{S_{k-2}}$. The receiver receives information for $\Sigma_{S_{k-1}}$ (as opposed

to Σ_{S_k}, despite the fact that it knows S_k; this is because the transmitter can only send for $\Sigma_{S_{k-1}}$), and sends for $\Sigma_{S_{k-1}}$. Again, the decoupling may be seen more easily by studying the control setup (e.g. (28) and Lemma 1).

G. Assumptions

In what follows, the choice of parameters according to (19)-(22) and Assumptions A1 and A2 are assumed unless otherwise specified:

A1: $\gamma(s[j]) > 0$ for each j, i.e., each state $s[j]$ is assigned with a nonzero power;

A2: $s[j] \neq 0$ for each j, i.e., the channel contains no erasure.

As a consequence of the assumptions, it holds that $a(s[j], s[l]) > 1$ for each j and l, that $\bar{a}[j] > 1$ for each j, and that $\tilde{a} > 1$ (refer to (20), (29), and (30) for the definitions).

We adopt these assumptions only for convenience; our main results remain true even if the assumptions do not hold. In fact, when A1 and A2 hold, we have $a(s[j], s[l]) > 1$ and that the mapping from $x_{k-1}^{(j)}$ to $x_k^{(j)}$ in (26) is always "strictly contractive", which simplifies the development of a convergence result (Lemma 2) that will be used to prove the main results. On the other hand, if A1 or A2 does not hold, then when the subsystem assigned with zero power is activated, or when the channel state is an erasure, the receiver receives only an AWGN, and *no* information about the to-be-transmitted message flows across either the forward channel or the feedback channel. At such a moment, the transmitter state, receiver state, and the receiver estimate remain their values, namely, we have $A(S_{k-2}, S_{k-1}) = I$ and $b(S_{k-2}, S_{k-1}) = 0$ for some k and hence $x_k = x_{k-1}$, which is not "strictly contractive". This would require a couple of extra steps in establishing the (same) main results in a few places of our development. For convenience, we would like to develop the main results under A1 and A2 in the main body of the paper, and defer the description of the extra steps until Appendix F.

We also note that A2 can be directly verified from the given channel model, and A1 can also be easily verified by 1) checking the optimal solution of $\gamma(s[i])$ computed efficiently by a numerical solver; or 2) applying the "complementary slackness" condition (an inequality constraint holds strictly if and only if its multiplier is zero, namely $\Gamma(s[i]) > 0$ if and only if the ith multiplier is zero; cf. [43]) to the optimization problem, which has been widely used to determine if a constraint is active or not.

H. Encoding/decoding method

Encoding and decoding

Fix the coding length to be $(k+1)$, i.e., we use the channel from time 0 to time k. We define $\mathcal{B} \in \mathbb{R}^m$ to be the unit hypercube centered at the origin, with each side (jth side denoted as $\mathcal{B}^{(j)}$) being the segment $[-\frac{1}{2}, \frac{1}{2}]$. For any fixed $\epsilon > 0$, and for each j, let $\mathcal{B}^{(j)}$ be uniformly partitioned into $\lfloor M_k^{(j)} \rfloor$ sub-intervals, where

$$M_k^{(j)} := \bar{a}[j]^{(k+1)(1-\epsilon)} \tag{40}$$

and $\lfloor M \rfloor$ denotes the largest integer no greater than M. For each k and j, it holds that

$$\lfloor M_k^{(j)} \rfloor = \frac{M_k^{(j)}}{\xi_k^{(j)}} \tag{41}$$

for some $\xi_k^{(j)} \in [1, 2)$ (noting that $M_k^{(j)} > 1$). Now $\mathcal{B}$ is partitioned into

$$M_k := \prod_{j=1}^{m} \lfloor M_k^{(j)} \rfloor \tag{42}$$

sub-hypercubes. Let the center of each sub-hypercube represent one of a set of M_k equally likely messages. Call the sub-hypercube centers the *codewords*, the sub-interval centers the *sub-codewords*, and the set of codewords the *codebook*.

For encoding, choose one codeword from the M_k centers, say W, and let $\boldsymbol{x}_0 := W$. Then $\boldsymbol{x}_0$ enters system (18) as the initial condition and generates channel input sequence u_0^k. For decoding, based on the channel output sequence y_0^k, the receiver calculates $\hat{x}_{0,k}^{(j)}$ for each j, and then decides the sub-interval center closest to $\hat{\boldsymbol{x}}_{0,k}$ to be the codeword transmitted by the transmitter. The decision at time k is denoted as $\hat{W}_k$. See Fig. 6 for a simple example of a codebook.
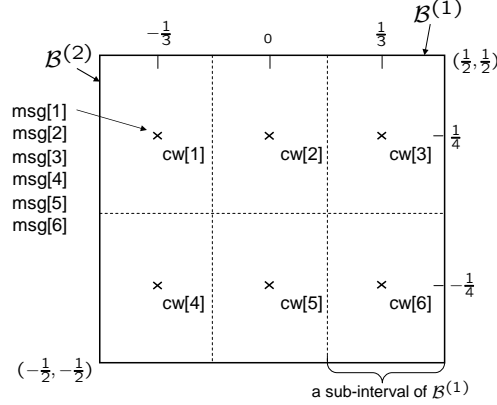

Fig. 6. An example of codebook. Assume $m = 2$, $\lfloor M_k^{(1)} \rfloor = 3$, and $\lfloor M_k^{(2)} \rfloor = 2$, namely $M_k = 6$. The six messages are represented by six codewords, the centers of the sub-hypercubes. Suppose that message msg[1] is to be conveyed. Then codeword cw[1] is to be transmitted, and the sub-codewords are $x_0^{(1)} = -1/3$ and $x_0^{(2)} = 1/4$. The two sub-codewords are transmitted through two decoupled subsystems. At the receiver side, if *both* sub-codewords are correctly decoded, then the codeword cw[1] and hence message msg[1] can be correctly recovered.

We see that the encoding and decoding are fairly simple. In fact, the computation complexity for the encoding and decoding mainly involves computing a product of $m \times m$ diagonal matrices (i.e. $\prod_{j=0}^{k} A(S_{j-1}, S_j)^{-1}$) and grows linearly in $(k+1)$, where $(k+1)$ is the number of channel uses.

Signalling rate

We define the signalling rate as

$$R := \lim_{k \to \infty} \frac{1}{k+1} \log M_k = \lim_{k \to \infty} \frac{1}{k+1} \sum_{j=1}^{m} \log \lfloor M_k^{(j)} \rfloor, \tag{43}$$

if the limit exists.

Probability of error

We declare a *decoding error* if the decoder's decision $\hat{W}_k$ is not equal to the transmitted codeword W. To compute the probability of error PE_k, we first compute the probability of error for the jth sub-codeword conditioned on the channel state sequence S_0^k:

$$PE_{k|S}^{(j)} := \Pr\left(\hat{x}_{0,k}^{(j)} \text{ and } x_0^{(j)} \text{ are in different sub-intervals of } \mathcal{B}^{(j)} | S_0^k\right). \tag{44}$$

Since conditioned on S_0^k, $x^{(i)}$ and $x^{(j)}$ evolve independently, we can independently compute $PE_{k|S}^{(j)}$ for each j (see Lemma 1 and Lemma 2 iii)). We then have that, $PE_{k|S}$, the probability of error for the codeword W conditioned on the channel state sequence S_0^k, is given by

$$PE_{k|S} = 1 - \prod_{j=1}^{m} (1 - PE_{k|S}^{(j)}). \tag{45}$$

16

Finally, the probability of error for decoding W, which averages $PE_{k|S}$ over all possible channel state sequences S_0^k, is

$$PE_k := \sum_{S_0^k \in \Omega_k} PE_{k|S} \Pr(S_0^k); \tag{46}$$

recall that Ω_k is defined as the set of all possible channel state sequences of length $(k+1)$. We remark that, though the above definitions are for some fixed $\boldsymbol{x}_0$, since the probability of error for different $\boldsymbol{x}_0$ shares the same asymptotic behavior [8], it is sufficient to study the probability of error for one fixed initial condition $\boldsymbol{x}_0$.

Power constraint and achievable rate

For any fixed k and fixed channel state sequence S_0^{k-1}, denote the channel inputs generated at time t as $u_{t|S_0^{t-1}}$ or simply $u_{t|S}$, $t = 0, 1, \cdots, k$, to emphasize the dependence on a specific channel state sequence (noticing that u_t does not specify how it depends on S_t). Then the average transmission power for this channel state sequence is

$$P_{k|S} := \frac{1}{k+1} \sum_{t=0}^{k} \mathbf{E}(u_{t|S})^2. \tag{47}$$

This corresponds to the average power for the transmission of one message for one possible channel state sequence. Then the transmission power averaged over all channel state sequences $S_0^k \in \Omega_{k-1}$ is

$$P_k := \sum_{S_0^{k-1} \in \Omega_{k-1}} P_{k|S} \Pr(S_0^{k-1}). \tag{48}$$

We say that the power constraint is satisfied if

$$\limsup_{k \to \infty} P_k \leq \mathcal{P} \tag{49}$$

for a given $\mathcal{P} > 0$.

We call a signalling rate R, defined in (43), *achievable* if, as k tends to infinity, PE_k decays to zero and the power constraint (49) is satisfied.

To conclude this section, we present Table I comparing the proposed scheme with other related communication schemes. In this table, a, b, and c are chosen according to (10). TRCSI denotes instantaneous, perfect transmitter-side and receiver-side CSI. Note that Elia's scheme in [12], Goldsmith and Varaiya's schemes in [2], and Viswanathan's schemes in [4] can be used for more general setups that are not consider in this table or in this paper. Also note that the original SK scheme's messages are chosen from the sub-interval center of interval $[0, 1]$, as opposed to interval $[-0.5, 0.5]$.

---

[8]In fact for k sufficiently large, $\hat{\boldsymbol{x}}_{0,k}$ has the form $\hat{\boldsymbol{x}}_{0,k} = \boldsymbol{x}_0 + \Delta_k$ (see (38)), where Δ_k does not depend on $\boldsymbol{x}_0$. Therefore, asymptotically the decoding error does not depend on $\boldsymbol{x}_0$.

TABLE I

COMPARISON OF THE PROPOSED SCHEME WITH OTHER COMMUNICATION SCHEMES

	Channel Model	Capacity	Tx Operations
Proposed	$y_k = S_k u_k + N_k$ DTRCSI, output feedback	$\frac{1}{2}\mathbf{E}_{S_{k-1}\sim\boldsymbol{\pi},S_k}\log(1+(S_k)^2\gamma(S_{k-1}))$	$\boldsymbol{x}_k = A(S_{k-2},S_{k-1})\boldsymbol{x}_{k-1} - \boldsymbol{b}(S_{k-2},S_{k-1})y_{k-1}$ $u_k = \boldsymbol{c}(S_{k-1})'\boldsymbol{x}_k$ update the transmitter state associated with S_{k-2}; transmit the transmitter state associated with S_{k-1}
SK's	$y_k = u_k + N_k$ output feedback	$\frac{1}{2}\log(1+\mathcal{P})$	$u_0 = a(0.5 - x_0)$ $k>0: u_k = a^k(\hat{x}_{0,k-1} - x_0)$
Gallager's	$y_k = u_k + N_k$ output feedback	$\frac{1}{2}\log(1+\mathcal{P})$	$x_k = a(x_{k-1} - v_{k-1})$ $u_k = x_k$
Elia's	$y_k = u_k + N_k$ output feedback	$\frac{1}{2}\log(1+\mathcal{P})$	$x_k = ax_{k-1}$ $u_k = cx_k - v_{k-1}$
Goldsmith & Varaiya's	$y_k = S_k u_k + N_k$ TRCSI	$\frac{1}{2}\mathbf{E}_{S_k\sim\boldsymbol{\pi}}\log(1+(S_k)^2\gamma(S_k))$	multiplexing according to S_k
Viswanathan's	$y_k = S_k u_k + N_k$ DTRCSI, with/without output feedback	$\frac{1}{2}\mathbf{E}_{S_{k-1}\sim\boldsymbol{\pi},S_k}\log(1+(S_k)^2\gamma(S_{k-1}))$	multiplexing according to S_{k-1}

	Rx Operations	Feedback Signal Available at Tx	Advantages	Disadvantages
Proposed	$\hat{\boldsymbol{x}}_{0,k} = \hat{\boldsymbol{x}}_{0,k-1} + \prod_{j=0}^{k} A(S_{j-1},S_j)^{-1}\boldsymbol{b}(S_{k-1},S_k)y_k$ demultiplex according to S_{k-1}	y_{k-1}, S_{k-1}	Bounded signals/parameters; unified operations for all k; control-oriented analysis readily applicable; same operations for digital/analog transmissions	Rx estimate of x_0 is biased (asym. unbiased)
SK's	$\hat{x}_{0,0} = 0.5 - a^{-1}y_0$ $k>0: \hat{x}_{0,k} = \hat{x}_{0,k-1} - a^{-k-2}\sqrt{a^2-1}\,y_k$	$\hat{x}_{0,k-1}$	unbiased Rx estimate of x_0; same operations for digital/analog transmissions	unbounded parameter a^k; initial operation differs from later ones
Gallager's	$\hat{x}_k = \hat{x}_{k-1} + a^{-k-1}by_k$ $\hat{x}_{0,k} = (1 - a^{-2k-2})^{-1}\hat{x}_k$	$v_{k-1} = a^{-1}by_{k-1}$	Bounded signals/parameters; unified operations for all k; unbiased Rx estimate of x_0	different operations for digital and analog transmissions
Elia's	$\hat{x}_k = a\hat{x}_{k-1} + by_k$ $\hat{x}_{0,k} = a^{-k}\hat{x}_k$	$v_{k-1} = cx_{k-1}$	unified operations for all k; control-oriented analysis readily applicable; same operations for digital/analog transmissions	unbounded feedback power; Rx estimate of x_0 is biased (asym. unbiased)
Goldsmith & Varaiya's	demultiplex according to S_k	S_k		
Viswanathan's	demultiplex according to S_{k-1}	S_{k-1} (y_{k-1})		output feedback not fully explored

IV. Achieving capacity of channel $\boldsymbol{F}$

In this section, we show that the feedback communication scheme proposed in Section III along with the parameters given by (19)-(22) is capacity-achieving. Our main result is summarized in Theorem 1.

Theorem 1. *Suppose that for channel $\boldsymbol{F}$, the channel state S_k forms an ergodic Markov process and is available instantaneously to the receiver and with one step delay to the transmitter. Given any $\mathcal{P} > 0$, let $\gamma(\cdot)$ be the capacity-achieving power allocation that maps the channel state S_k to the channel input power $\gamma(S_k)$ and such that $\mathbf{E}_{S\sim\boldsymbol{\pi}}\gamma(S) \leq \mathcal{P}$ holds. Then, the feedback communication scheme described in Section III, along with the parameters given by (19)-(22) under Assumptions A1 and A2, achieves any rate arbitrarily close to the feedback capacity*

$$C = \frac{1}{2}\mathbf{E}_{S_k\sim\boldsymbol{\pi},S_{k+1}}\log(1 + (S_{k+1})^2\gamma(S_k)) = \log\tilde{a} \tag{50}$$

under average input power constraint

$$\mathbf{E}u^2 \leq \mathcal{P}. \tag{51}$$

To prove this theorem, we first compute the achievable rate, followed by showing that the power constraint (49) is satisfied. See Appendix A for the proof of $C = \log\tilde{a}$.

We note that in [4], the capacity for Gaussian channels as given in (50) is only in terms of the maximum mutual information. Whether it is equal to the operational capacity is not explicitly proven in [4]. ([4] proved the achievability and converse for discrete FSMCs but not for Gaussian FSMCs.) To be complete, we provide the converse in Appendix B. Combining the converse and the achievability shown in Theorem 1, we establish that the maximum mutual information is indeed equal to the operational capacity.

A. Achievable rate

In this part, we prove that for any $\epsilon > 0$ small enough, rate $(1-\epsilon)C$ is achievable. This development is facilitated by considering the control setup. In fact, for an AFSMC, whenever the control setup in Fig. 4 (a) is open-loop unstable but closed-loop stabilized, the communication system in Fig. 3 can achieve any rate smaller than the growth rate of the open-loop control setup, similar to the case of Gaussian channels without time-selective fading [12]. Thus, our first step is to establish relevant stability results for the control setup.

Our choice of parameters in Section III-D leads to a conclusion that the open loop of the control setup is unstable. The average rate of growth for this unstable system is $\log\tilde{a}$. This is exactly the supremum achievable rate for the proposed communication system. Next, we show that the control setup is stabilized in closed-loop, based on which we prove the signalling rate is achievable. Note that the stability of the control setup is in the sense of, firstly, the boundedness and convergence of the first moment $\mathbf{E}\boldsymbol{x}_k$ to zero, and secondly, the boundedness of the variance-covariance matrix $\Sigma_k := \mathbf{E}\boldsymbol{x}_k\boldsymbol{x}'_k - \mathbf{E}\boldsymbol{x}_k\mathbf{E}\boldsymbol{x}'_k$, *both conditioned on a given channel state sequence $\{S_k\}$ and initial condition $\boldsymbol{x}_0$.* (Note that when conditioned on $\{S_k\}$ and $\boldsymbol{x}_0$, the randomness in $\boldsymbol{x}_k$ comes from AWGN N_0^{k-1}.) Finally, note that if one fixes the choice of $A(S_k, S_{k+1})$, then the open-loop growth rate as well as the supremum signalling rate is fixed, and thus $\boldsymbol{b}(S_k, S_{k+1})$ can be chosen to ensure the stability of the closed-loop as well as the minimum average input power (in order to achieve the most power-efficient communication), which is exactly a cheap control problem.

Lemma 2. *Assume the hypotheses of Theorem 1, and fix a channel state sequence $\{S_k\}$ in Ω. Then for the control setup (23),*

i) Φ_k *is the state transition matrix, namely the response due to initial condition $\boldsymbol{x}_0$ is $\boldsymbol{x}_{k+1} = \Phi_k \boldsymbol{x}_0$. For every $j = 1, \cdots, m$, it holds that $0 < \phi_k^{(j)} \leq 1$ for any k and*

$$\phi_k^{(j)} \xrightarrow{\text{P}} 0 \tag{52}$$

for any j;

ii) For any fixed initial condition $\boldsymbol{x}_0$, it holds that for any j, $-|x_0^{(j)}| \leq \mathbf{E} x_k^{(j)} \leq |x_0^{(j)}|$ for any k and that $\mathbf{E} x_k^{(j)} \xrightarrow{\text{P}} 0$;

iii) For any $\boldsymbol{x}_0$, $\Sigma_k := \mathbf{E}(\boldsymbol{x}_k - \mathbf{E}\boldsymbol{x}_k)(\boldsymbol{x}_k - \mathbf{E}\boldsymbol{x}_k)'$ is a diagonal matrix, i.e., the components of $\boldsymbol{x}_k$ are mutually independent, i.e. $\mathbf{E} x_k^{(i)} x_k^{(j)} = \mathbf{E} x_k^{(i)} \mathbf{E} x_k^{(j)}$ if $i \neq j$, and there exist $\underline{c}$ and $\bar{c}$ such that

$$0 < \underline{c} \leq \Sigma_k^{(j)} \leq \bar{c} \leq \infty \tag{53}$$

for any k, j, and $\{S_k\}$, where $\Sigma_k^{(j)}$ is the (j,j)th element of Σ_k.

Proof: See Appendix C. $\qquad\square$

Now we can use the stability of the control setup to establish the reliable communication across the channel. Note that the stability of the control setup implies that any signal in the communication scheme is bounded, as claimed in Section I.

Remark 1. The intuition behind the following development is that, in view of the stability shown in Lemma 2, the difference between $x_0^{(j)}$ and $x_{0,k}^{(j)}$ vanishes sufficiently fast in both the first and second moments as time increases thanks to closed-loop stability conditions, and so does $PE_{k|S}^{(j)}$, the probability that $x_0^{(j)}$ and $x_{0,k}^{(j)}$ are in different sub-intervals, if the channel state sequence is typical. Then because of the decoupling proven in Lemma 1 and Lemma 2, $PE_{k|S}$ can be shown to be vanishing.

Proposition 1. *Assume the hypotheses of Theorem 1. Then the communication system reliably transmits at rate*

$$R = (1 - \epsilon) \log \tilde{a} = (1 - \epsilon) C, \tag{54}$$

for any given $\epsilon > 0$.

Proof: To establish the achievable rate, we first compute the signalling rate, followed by proving that the average probability of error PE_k goes to zero as k goes to infinity, which implies that the signalling rate is achievable.

Signalling rate

It holds that

$$
\begin{aligned}
R &:= \lim_{k \to \infty} \frac{\sum_{j=1}^m \log \lfloor M_k^{(j)} \rfloor}{k+1} \\
&\overset{(41)}{=} \lim_{k \to \infty} \left(\frac{\sum_{j=1}^m \log M_k^{(j)}}{k+1} - \frac{\sum_{j=1}^m \log \xi_k^{(j)}}{k+1} \right) \\
&\overset{(29)}{=} \lim_{k \to \infty} \frac{\sum_{j=1}^m \log \bar{a}[j]^{(k+1)(1-\epsilon)}}{k+1} \\
&= (1 - \epsilon) \sum_{j=1}^m \log \bar{a}[j] \\
&\overset{(30)}{=} (1 - \epsilon) \log \tilde{a}.
\end{aligned}
\tag{55}
$$

Probability of error

The proof of vanishing probability of error is essentially an AEP-based argument (cf. [36]). Fix $\epsilon > 0$. Fix $k > 0$. Recall the definition in (4) for the (k, μ)-typical set $\Omega_{k,\mu}$ where $\mu > 0$ is to be determined as a function of ϵ but independent of k. Then Ω_k is partitioned into two subsets: $\Omega_{k,\mu}$ and $\Omega_{k,\mu}^c$, where the superscript ^{c} denotes the complement in Ω_k. Our goal is to show that the probability of error $PE_{k|S}$ satisfies

$$PE_{k|S} \leq \kappa(k), \quad \forall S_0^k \in \Omega_{k,\mu} \tag{56}$$

for some vanishing function $\kappa(\cdot) \geq 0$ which is independent of k (i.e., $\kappa(\cdot)$ is a time-invariant function). This would lead to vanishing probability of error PE_k. More precisely, note that

$$
\begin{aligned}
PE_k \quad &:= \quad \sum_{S_0^k \in \Omega_k} PE_{k|S} \Pr(S_0^k) \\
&= \quad \sum_{S_0^k \in \Omega_{k,\mu}} PE_{k|S} \Pr(S_0^k) + \sum_{S_0^k \in \Omega_{k,\mu}^c} PE_{k|S} \Pr(S_0^k) \\
&\leq \quad \sum_{S_0^k \in \Omega_{k,\mu}} \kappa(k) \Pr(S_0^k) + \sum_{S_0^k \in \Omega_{k,\mu}^c} \Pr(S_0^k) \\
&\leq \quad \kappa(k) + \Pr(\Omega_{k,\mu}^c),
\end{aligned}
\tag{57}
$$

where we have used $\kappa(\cdot) \geq 0$, (56), $\sum_{S_0^k \in \Omega_{k,\mu}} \Pr(S_0^k) \leq 1$, and $PE_{k|S} \leq 1$. Therefore, as k increases, the average probability of error PE_k would decay to zero.

Now we start by investigating the behavior of $PE_{k|S}^{(j)}$ on $\Omega_{k,\mu}$. Fix any channel state sequence S_0^k and initial condition $\boldsymbol{x}_0$. Because the noise is zero mean i.i.d. Gaussian, by Lemma 1, $\boldsymbol{x}_{k+1}$ is Gaussian with distribution $\mathcal{N}(\mathbf{E}\boldsymbol{x}_{k+1}, \Sigma_{k+1})$, *if conditioned on S_0^k and $\boldsymbol{x}_0$*, where Σ_{k+1} is defined in Lemma 2 iii). Then notice that $\mathbf{E}\boldsymbol{x}_{k+1} = \Phi_k \boldsymbol{x}_0$. Therefore, $\hat{\boldsymbol{x}}_{0,k}$ is an m-variate Gaussian vector distributed as

$$\mathcal{N}\left(\boldsymbol{x}_0 - (\Phi_k)^2 \boldsymbol{x}_0, (\Phi_k)^2 \Sigma_{k+1}\right), \tag{58}$$

and particularly, for each j, $\hat{x}_{0,k}^{(j)}$ is a univariate Gaussian distributed as

$$\mathcal{N}\left(x_0^{(j)} - (\phi_k^{(j)})^2 x_0^{(j)}, (\phi_k^{(j)})^2 \Sigma_{k+1}^{(j)}\right). \tag{59}$$

We now assume without loss of generality that $x_0^{(j)}$ is the center of the ith sub-interval of $\mathcal{B}^{(j)}$. See Fig. 7. We can thus derive the following expression of $PE_{k|S}^{(j)}$, which by Lemma 1 and Lemma 2 iii) can be computed independently for each j:

$$
\begin{aligned}
PE_{k|S}^{(j)} \quad &= \quad Q\left(\frac{0.5/\lfloor M_k^{(j)}\rfloor + x_0^{(j)}(\phi_k^{(j)})^2}{\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}}\right) + Q\left(\frac{0.5/\lfloor M_k^{(j)}\rfloor - x_0^{(j)}(\phi_k^{(j)})^2}{\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}}\right) \\
&= \quad Q\left(\frac{0.5\xi_k^{(j)}/M_k^{(j)} + x_0^{(j)}(\phi_k^{(j)})^2}{\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}}\right) + Q\left(\frac{0.5\xi_k^{(j)}/M_k^{(j)} - x_0^{(j)}(\phi_k^{(j)})^2}{\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}}\right) \\
&= \quad Q\left(\frac{0.5\xi_k^{(j)}}{\bar{a}[j]^{(k+1)(1-\epsilon)}\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}} + \frac{x_0^{(j)}\phi_k^{(j)}}{\sqrt{\Sigma_{k+1}^{(j)}}}\right) + Q\left(\frac{0.5\xi_k^{(j)}}{\bar{a}[j]^{(k+1)(1-\epsilon)}\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}} - \frac{x_0^{(j)}\phi_k^{(j)}}{\sqrt{\Sigma_{k+1}^{(j)}}}\right).
\end{aligned}
\tag{60}
$$

To see the convergence to zero of $PE_{k|S}^{(j)}$, note that by the uniform boundedness of $\xi_k^{(j)} \in [1,2)$, $x_0^{(j)} \in (-0.5, 0.5)$, $\phi_k^{(j)} \in (0,1]$, and $\Sigma_{k+1}^{(j)}$, it is sufficient to show that $\bar{a}[j]^{(k+1)(1-\epsilon)}\phi_k^{(j)}$ goes to zero. One can show that

$$
\begin{aligned}
\bar{a}[j]^{(k+1)(1-\epsilon)}\phi_k^{(j)} \quad &= \quad \exp\left[(k+1)\left((1-\epsilon)\log\bar{a}[j] + \frac{1}{k+1}\log\phi_k^{(j)}\right)\right] \\
&= \quad \exp\left[(k+1)\sum_{l=1}^m \left((1-\epsilon)\pi[j]p_{jl} - \frac{n(j,l,k)}{k+1}\right)\log a(s[j], s[l])\right] \\
&= \quad \exp\left[(k+1)\sum_{l=1}^m d(j,l,k)\log a(s[j], s[l])\right],
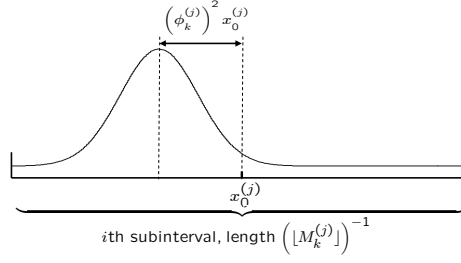\end{aligned}
\tag{61}
$$

Fig. 7. The location of $x_0^{(j)}$ in a sub-interval of $\mathcal{B}^{(j)}$ and the distribution of $\hat{x}_{0,k}^{(j)}$, with mean $\left(x_0^{(j)} - (\phi_k^{(j)})^2 x_0^{(j)}\right)$ and variance $(\phi_k^{(j)})^2 \Sigma_{k+1}^{(j)}$.

where

$$d(j,l,k) := (1-\epsilon)\pi[j]p_{jl} - \frac{n(j,l,k)}{k+1}. \tag{62}$$

If $p_{jl} = 0$ (noting that $\pi[j] \neq 0$ by ergodicity), then $d(j,l,k) \leq 0$ for any S_0^k in $\Omega_{k,\mu}$. If, however, $p_{jl} \neq 0$, then let

$$
\begin{aligned}
\frac{n(j,l,k)}{n(j,k)p_{jl}} &= 1 + \lambda_1 \frac{\mu}{p_{jl}} \\
\frac{n(j,k)}{(k+1)\pi[j]} &= 1 + \lambda_2 \frac{\mu}{\pi[j]},
\end{aligned}
\tag{63}
$$

where λ_1 and λ_2 are such that $|\lambda_1| \leq 1$ and $|\lambda_2| \leq 1$ by (4). Then it holds that

$$
\begin{aligned}
1 - \epsilon - \frac{n(j,l,k)}{n(j,k)p_{jl}}\frac{n(j,k)}{(k+1)\pi[j]} &= -\epsilon - \mu\left(\frac{\lambda_1}{p_{jl}} + \frac{\lambda_2}{\pi[j]} + \frac{\mu\lambda_1\lambda_2}{p_{jl}\pi[j]}\right) \\
&\leq -\epsilon + \mu\left(\frac{1}{p_{jl}} + \frac{1}{\pi[j]} + \frac{\mu}{p_{jl}\pi[j]}\right).
\end{aligned}
\tag{64}
$$

Therefore, for sufficiently small $\mu_1 > 0$ (more specifically, $\mu_1 := \mu_1(\epsilon,j,l)$ depends on ϵ, j, and l but independent of k), $d(j,l,k)$ is non-positive for any S_0^k in Ω_{k,μ_1}. Furthermore, as the choices of j and l are finite, there exists a sufficiently small $\mu_2 > 0$ ($\mu_2 := \mu_2(\epsilon)$ depends on ϵ but independent of k) such that $d(j,l,k)$ is non-positive for any S_0^k in Ω_{k,μ_2}.

Let $a := \min_l a(s[j], s[l]) > 1$ and $\alpha := a^\epsilon > 1$. It then follows from (61), (62), and $d(j,l,k) \leq 0$ that

$$
\begin{aligned}
\bar{a}[j]^{(k+1)(1-\epsilon)}\phi_k^{(j)} &\leq a^{(k+1)\sum_{l=1}^m d(j,l,k)} \\
&= a^{(k+1)(1-\epsilon)\pi[j]-n(j,k)} \\
&= a^{-\epsilon n(j,k)+(k+1)(1-\epsilon)\left(\pi[j]-\frac{n(j,k)}{k+1}\right)} \\
&= \alpha^{-n(j,k)+(k+1)\left(\frac{1}{\epsilon}-1\right)\left(\pi[j]-\frac{n(j,k)}{k+1}\right)} \\
&:= \alpha^{-\eta_k^{(j)}},
\end{aligned}
\tag{65}
$$

where we have defined

$$\eta_k^{(j)} := n(j,k) - (k+1)\left(\frac{1}{\epsilon}-1\right)\left(\pi[j] - \frac{n(j,k)}{k+1}\right). \tag{66}$$

Therefore,

$$
\begin{aligned}
Q_{k,1} &:= Q\left(\frac{0.5\xi_k^{(j)}}{\bar{a}[j]^{(k+1)(1-\epsilon)}\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}} + \frac{x_0^{(j)}\phi_k^{(j)}}{\sqrt{\Sigma_{k+1}^{(j)}}}\right) \\
&\leq Q\left(\frac{\xi_k^{(j)}}{2\sqrt{\Sigma_{k+1}^{(j)}}}\alpha^{n(j,k)-(k+1)\left(\frac{1}{\epsilon}-1\right)\left(\pi[j]-\frac{n(j,k)}{k+1}\right)} + \frac{x_0^{(j)}\phi_k^{(j)}}{\sqrt{\Sigma_{k+1}^{(j)}}}\right) = Q\left(\alpha^{\eta_k^{(j)}+\zeta_k^{(j)}}\right),
\end{aligned}
$$

22

where

$$\zeta_k^{(j)} := \log_\alpha \left\{ \frac{\xi_k^{(j)} + 2\alpha^{-n(j,k)\left[1-\left(\frac{1}{\epsilon}-1\right)\left(\frac{(k+1)\pi[j]}{n(j,k)}-1\right)\right]} x_0^{(j)} \phi_k^{(j)}}{2\sqrt{\Sigma_{k+1}^{(j)}}} \right\}. \tag{67}$$

Similar to the way that we chose μ_2, we may choose $\mu_3 := \mu_3(\epsilon)$ sufficiently small such that for any $S_0^k \in \Omega_{k,\mu_3}$ and for any j and k, it holds that

$$1 - \left(\frac{1}{\epsilon} - 1\right)\left(\frac{(k+1)\pi[j]}{n(j,k)} - 1\right) > 0; \tag{68}$$

in fact, if

$$\pi[j] - \frac{\mu_3}{\epsilon} > \sigma_1 \tag{69}$$

holds for some $\mu_3 > 0$ and $\sigma_1 > 0$ and for all $j = 1, 2, \cdots, m$, then (68) is true. Here σ_1 depends on ϵ but does not depend on k. Note that such μ_3 and σ_1 exist since m is finite. Moreover, by $\xi_k^{(j)} \in [1, 2)$, $x_0^{(j)} \in (-0.5, 0.5)$, $\phi_k^{(j)} \in (0, 1]$, and $\Sigma_{k+1}^{(j)} \in (\underline{c}, \bar{c})$ for any j, k, and S_0^k (by Lemma 2), one can easily show that ζ_k is uniformly bounded, namely $\zeta_k \in (\underline{\zeta}_1, \bar{\zeta}_1)$ where the bounds are independent of j, k, $\boldsymbol{x}_0$, and S_0^k.

Now pick $\mu_4 > 0$ sufficiently small such that $\mu_4 \leq \min\{\mu_2, \mu_3\}$. Consequently, for any $S_0^k \in \Omega_{k,\mu_4}$, it holds that

$$\left(\frac{1}{\epsilon} - 1\right)\left|\pi[j] - \frac{n(j,k)}{k+1}\right| < \left(\frac{1}{\epsilon} - 1\right)\mu_4. \tag{70}$$

This yields that,

$$\begin{aligned} \eta_k &\geq (k+1)\left(\frac{n(j,k)}{k+1} - \left(\frac{1}{\epsilon}-1\right)\mu_4\right) \\ &= (k+1)\left(\pi[j] + \lambda\mu_4 - \left(\frac{1}{\epsilon}-1\right)\mu_4\right) \\ &\geq (k+1)\left(\pi[j] - \frac{\mu_4}{\epsilon}\right) \\ &> (k+1)\sigma, \end{aligned} \tag{71}$$

where $|\lambda| \leq 1$ and $\sigma > 0$. Thus, $(\eta_k + \zeta_k)$ goes to infinity and $Q_{k,1}$ vanishes. More precisely,

$$Q_{k,1} < Q(\alpha^{(k+1)\sigma + \underline{\zeta}_1}). \tag{72}$$

Similarly,

$$\begin{aligned} Q_{k,2} &:= Q\left(\frac{0.5\xi_k^{(j)}}{\bar{a}[j]^{(k+1)(1-\epsilon)}\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}} - \frac{x_0^{(j)}\phi_k^{(j)}}{\sqrt{\Sigma_{k+1}^{(j)}}}\right) \\ &< Q(\alpha^{(k+1)\sigma + \underline{\zeta}_2}). \end{aligned} \tag{73}$$

for any $S_0^k \in \Omega_{k,\mu_5}$ and for some $\underline{\zeta}_2$, where $\mu_5 > 0$.

Letting $\mu_\cdot = \min\{\mu_4, \mu_5\}$ and $\underline{\zeta} := \min\{\underline{\zeta}_1, \underline{\zeta}_2\}$, we have that

$$PE_{k|S}^{(j)} = Q_{k,1} + Q_{k,2} < 2Q(\alpha^{(k+1)\sigma + \underline{\zeta}}), \tag{74}$$

which decays to zero as k tends to infinity for any $S_0^k \in \Omega_{k,\mu}$. Then invoking the union bound

$$PE_{k|S} = 1 - \prod_{j=1}^m (1 - PE_{k|S}^{(j)}) \leq \sum_{j=1}^m PE_{k|S}^{(j)} < 2mQ(\alpha^{(k+1)\sigma + \underline{\zeta}}), \tag{75}$$

we conclude that $PE_{k|S}$ would converge to zero on $\Omega_{k,\mu}$ for sufficiently small μ. Thus we prove that PE_k decays to zero, i.e., rate R is achievable. [9] $\qquad\square$

---

[9]We may also employ a modified decoding after we obtain $\hat{\boldsymbol{x}}_{0,k}$, by letting $\hat{W}_k$ be the sub-interval center closest to $(I - (\Phi_k)^2)^{-1}\hat{\boldsymbol{x}}_{0,k}$. This removes the estimate bias. The asymptotic behavior analysis of the communication scheme remains the same.

B. Power computation

Proposition 2. *Assume the hypotheses of Theorem 1. Then the average channel input power is*

$$P := \limsup_{k \to \infty} P_k = \sum_{j=1}^{m} \pi[j]\gamma(s[j]), \tag{76}$$

and hence satisfies power constraint (49).

Remark 2. The idea of the proof is as follows. We show that the average power of the ith subsystem converges to $\gamma(s[i])$, and the ith subsystem is selected to generate the channel input if and only if the *previous* fade was $s[i]$. Hence, the average power used by the communication scheme (18) converges to the weighted sum of $\gamma(s[i])$, the optimal power. Note that the channel input u_k depends on S_{k-1}, and hence the channel input power at each time depends on the previous channel state. This does not contradict (20), in which $\gamma(S_{k-2})$ is used for generating x_k. This is because, by design, that the effect of $\gamma(S_{k-2})$ would not be reflected in the channel input at time $(k-1)$, the immediate next step, but at time $t > (k-1)$ such that $S_{t-1} = S_{k-2}$. In other words, the channel input power at t depends on S_{k-2}, or effectively depends on S_{t-1}.

Proof: See Appendix D. □

Combining Propositions 1 and 2, we have completed the proof for Theorem 1.

C. Further performance analysis: Doubly exponential decay for typical state sequences

In this subsection, we show that the proposed feedback communication scheme leads to a doubly exponential decay of the probability of error when the channel state behavior is "typical", as defined below.

Definition 3. *Define the* typical set *of sequences* $\{S_k\}$ *as*

$$\Omega_{TYP} := \left\{ \{S_k\} \,\middle|\, \frac{n(j,k)}{k+1} \to \pi[j] \text{ and } \frac{n(j,l,k)}{n(j,k)} \to p_{jl} \text{ as } k \to \infty \right\}. \tag{77}$$

Each sequence in Ω_{TYP} *is called a* typical state sequence.

By ergodicity of $\{S_k\}$, it holds that $\Pr(\Omega_{TYP}) = 1$. Hereafter, by "with probability one" or "almost surely" or "a.s." we mean "for every channel state sequence $\{S_k\} \in \Omega_{TYP}$" or "for every typical state sequence". Then by the above definition, it holds that

$$\begin{aligned}
\frac{n(j,k)}{k+1} &\overset{\text{a.s.}}{\to} \pi[j] \\
\frac{n(j,l,k)}{n(j,k)} &\overset{\text{a.s.}}{\to} p_{jl} \\
\frac{n(j,l,k)}{k+1} &\overset{\text{a.s.}}{\to} \pi[j]p_{jl}
\end{aligned} \tag{78}$$

as k tends to infinity.

We then have

Proposition 3. *Assume the hypotheses of Theorem 1. Then*
i) For sufficiently large k, for any j and $\{S_k\} \in \Omega_{TYP}$,

$$PE_{k|S}^{(j)} \le \beta_1 \exp\left[-\exp((k+1)\beta_2 + o(k+1))\right] \tag{79}$$

for some $\beta_1, \beta_2 > 0$, where $o(k)$ denotes the terms with lower order than k, namely $o(k)/k$ vanishes as k tends to infinity;

ii) For any j and $\{S_k\} \in \Omega_{TYP}$, the decay exponent for $PE_{k|S}^{(j)}$ is

$$e^{(j)} := \lim_{k \to \infty} \frac{1}{k} \log \left(\log \left(\frac{1}{PE_{k|S}^{(j)}} \right) \right) = 2\epsilon \log(\bar{a}[j]), \tag{80}$$

and the decay exponent for $PE_{k|S}$ is

$$e := \lim_{k \to \infty} \frac{1}{k} \log \left(\log \left(\frac{1}{PE_{k|S}} \right) \right) = \min_j e^{(j)} = 2\epsilon \log \underline{a}, \tag{81}$$

where $\underline{a} := \min_j \bar{a}[j]$.

In other words, this proposition claims that, for each sub-codeword $x_0^{(j)}$, essentially its probability of error decays doubly exponentially, and hence $PE_{k|S}$ decays doubly exponentially with respect to k for large enough k, provided that we exclude a set of $\{S_k\}$ that would occur with zero probability and on which the signalling rate cannot be achieved [10]. The decay exponent of $PE_{k|S}$ is given by the smallest (slowest) decay exponent of $PE_{k|S}^{(j)}$ among all j. When the channel has no fading, i.e., $m = 1$ and $a = a(s[j], s[l])$ for all j and l, the decay exponent becomes $e = 2\epsilon \log a = 2(C - R)$, and we recover the decay exponent obtained in [10], [11].

Remark 3. To be more precise, this double-exponential decay argument builds on an infinite-length channel state realization sequence $\{S_k\}$. Then there corresponds to each fixed coding length $(k + 1)$ a communication problem in which the channel state sequence is S_0^k, a prefix of $\{S_k\}$. Each problem has its own message (one out of the M_k possible messages) to be transmitted, and each problem generates its own AWGN [11]. A probability of error PE_k is then obtained whose upper bound vanishes doubly exponentially as k goes to infinity, if $\{S_k\}$ is typical. Since typical state sequences form a probability one set, we conclude that with probability one, $PE_{k|S}$ decays doubly exponentially; in other words, almost every "sample trajectory" of $PE_{k|S}$ decays doubly exponentially. It is also worth noting that the *average* probability of error PE_k (see (46)) decays only singly exponentially, though $PE_{k|S}$ decays doubly exponentially with probability one. In fact, this proposition concerning $PE_{k|S}$ is based on the Strong Law of Large Numbers whereas the decay of PE_k is based on the Weak Law of Large Numbers. The Strong Law (and hence the doubly exponential statement for $PE_{k|S}$ over Ω_{TYP}) holds for the set of channel state sequences $\{S_k\}$ of *infinite* lengths, whereas the Weak Law (and hence the singly exponential statement for PE_k over Ω_k) holds for $\{\Omega_k\}_{k=0}^{\infty}$, the sequence of sets Ω_k consisting of channel state sequences S_0^k of *finite* lengths (though k can be very large, it is not infinity).

Remark 4. Despite the fact that all sequences $\{S_k\} \in \Omega_{TYP}$ have the same decay exponent, the convergence to the decay exponent is not uniform in k and not uniform over Ω_{TYP}. This is because for any given sequence, how far $(1/k) \log(\log(1/PE_{k|S}))$ is away from the decay exponent depends on the $(k + 1)$-length prefix of the sequence, which varies for each k and from sequence to sequence.

Remark 5. Similar to the AWGN case, the doubly exponential decay is possible if an *average* channel input power constraint is used (cf. [44]); a singly exponential decay is expected if a peak power constraint is used.

Proof: See Appendix E. □

---

[10] If the channel state sequence is not typical, then there exists at least one channel state $s[j]$ whose subsystem uses the forward channel and feedback channel less often than it typically does, resulting in possibly insufficient refinement of sub-codeword $x_0^{(j)}$, i.e., $\hat{x}_{0,k}^{(j)}$ may not be close enough to $x_0^{(j)}$. Thus, the transmitter codeword may not be correctly decoded. In other words, the probability of error does not decay to zero and the signalling rate is not achievable.

[11] The message and the noise in the problem for $(k + 1)$ are not necessarily nested in the problem for $(k + 2)$, as the upper bound of the probability of error does not depend on which message is selected or on the realizations of the noise.

D. Transmission of Gaussian random vectors

With some modifications we can transmit a Gaussian random vector over channel $\boldsymbol{F}$, in parallel with the AWGN channel case. Let us use the same parameters given in (19)-(22) and $\boldsymbol{x}_0 \sim \mathcal{N}(0, \mathcal{P}I_m)$, and follow the dynamics (18). For a given channel state sequence $\{S_k\} \in \Omega_{TYP}$, we obtain the MSE distortion as

$$\text{MSE}(\hat{\boldsymbol{x}}_{0,k}) := \mathbf{E}(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0,k})(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0,k})' = (\Phi_k)^2 \mathbf{E}\boldsymbol{x}_{k+1}\boldsymbol{x}'_{k+1}, \tag{82}$$

which, by rate-distortion theory, requires an asymptotic rate to be at least

$$
\begin{aligned}
\lim_{k\to\infty} \frac{1}{2(k+1)} \log \frac{\mathcal{P}^m}{\det \text{MSE}(\hat{\boldsymbol{x}}_{0,k})} &= \lim_{k\to\infty} \frac{1}{2(k+1)} \log \frac{\mathcal{P}^m}{\prod_{j=1}^m (\det(\mathbf{E}\boldsymbol{x}_{k+1}\boldsymbol{x}'_{k+1})\phi_k^{(j)})^2} \\
&\overset{(a)}{=} -\lim_{k\to\infty} \frac{1}{k+1} \prod_{j=1}^m \log \phi_k^{(j)} \\
&\overset{(b)}{=} \log \tilde{a},
\end{aligned}
$$

where (a) follows from the boundedness of $\mathcal{P}^m$ and $\det(\mathbf{E}\boldsymbol{x}_{k+1}\boldsymbol{x}'_{k+1})$, and (b) follows from a derivation similar to Appendix A. Since $\log \tilde{a}$ is the capacity, we conclude that $\boldsymbol{x}_0$ is successively refined at the capacity rate of channel $\boldsymbol{F}$.

E. Special case: AWGN i.i.d. fading channel

Theorem 1 directly applies to the case that $\{S_k\}$ forms a discrete i.i.d. process. However, a simplified capacity-achieving feedback scheme with a *scalar* transmitter state exists. Assume that the channel state has an i.i.d. distribution given by

$$\Pr(S_k = s[i]) = p[i] \text{ for } k = 0, 1, \cdots, \tag{83}$$

where for $i = 1, 2, \cdots, m$, $p[i]$ and $s[i]$ are fixed numbers. Given any power budget $\mathcal{P} > 0$, we choose the parameters in the communication scheme as

$$
\begin{aligned}
A(S_{k-2}, S_{k-1}) &:= A(S_{k-1}) &&:= \sqrt{(S_{k-1})^2 \mathcal{P} + 1} &&\in \mathbb{R} \\
\boldsymbol{b}(S_{k-2}, S_{k-1}) &:= \boldsymbol{b}(S_{k-1}) &&:= \frac{S_{k-1}\mathcal{P}}{\sqrt{(S_{k-1})^2 \mathcal{P} + 1}} &&\in \mathbb{R} \\
\boldsymbol{c}(S_{k-1}) &:= \boldsymbol{c} &&:= 1 &&\in \mathbb{R}.
\end{aligned} \tag{84}
$$

Note that A and $\boldsymbol{b}$ in this design do *not* require the augmented channel state (S_{k-2}, S_{k-1}); S_{k-1} is sufficient. As a consequence, no multiplexing is needed. The scalar dynamics of the associated control setup evolves according to

$$x_k = A(S_{k-1})^{-1} x_{k-1} - b(S_{k-1}) N_{k-1}, \tag{85}$$

where x_k is a scalar. We can show that this design leads to a transmission rate arbitrarily close to the capacity (proof skipped for brevity)

$$C_{iid} = \frac{1}{2} \sum_{i=1}^m p[i] \log(1 + s[i]^2 \mathcal{P}) = \frac{1}{2}\mathbf{E}_S \log(1 + S^2 \mathcal{P}). \tag{86}$$

Note that no power adaptation is present in the capacity formula and in the proposed scheme.

1) Extension: AWGN i.i.d. fading channel with infinite channel states: AWGN i.i.d. fading channels with infinite channel states include many channels as the special cases, such as the Rayleigh, Rician, Nakagami, and Weibull fading channels. Here we focus on the scenario of real channel state-spaces; the scenario of complex channel state-spaces can be studied likewise. Assume that the channel state forms an i.i.d. process with density $p_S(s)$ defined on $\mathbb{R}$ and that the first and second moments exist. Then the channel capacity is

$$C_{iid,inf} = \frac{1}{2}\mathbf{E}_{S \sim p_S} \log(1 + S^2\mathcal{P}). \tag{87}$$

Then we construct a coding scheme with a scalar transmitter state as in the finite state-space case, using the choice of parameters given in (84). As we show in Appendix G, this scheme achieves any rate below the feedback capacity given in (87). We point out that the proof makes use of the fact that the transmitter can be designed as a *scalar* system and hence this proof may not be directly applicable to Markov channels with infinite states.

V. NUMERICAL EXAMPLE

Consider a Gilbert-Elliot fading channel with AWGN, i.e. an AFSMC with only two states, as illustrated in Fig. 8. We simulate the proposed scheme for this channel. Fig. 9 (a) shows the simulated $PE_{k|S}^{(j)}$ and $PE_{k|S}$ for a randomly chosen $\{S_k\}$, as well as the theoretic $PE_{k|S}$ computed from (60) and (45), where $\epsilon > 0$ is a slack from the Shannon capacity C, i.e., the signalling rate is $R = (1-\epsilon)C$. We see that the probability of error decays rather fast within 20 channel uses. However, the decay of $PE_{k|S}^{(j)}$ and $PE_{k|S}$ is not quite smooth, caused by instantaneous deviations from the typical channel state behavior (namely, for some k, either $|n(j,k)/(k+1)-\pi[j]|$, or $|n(j,l,k)/n(j,k)-p_{jl}|$, or both, are not sufficiently close to zero), though $\{S_k\}$ may be typical; see also Remark 3. This may be improved by considering a "turbo mode" of using larger power at the moments with large instantaneous deviations from the typical state behavior, which does not affect the average power constraint (under further investigation); see [45] for the idea of the turbo mode. Fig. 9 (b) shows the decay of PE_k [12]. These fast decays imply that the proposed scheme allows shorter coding lengths and shorter coding delays; here the coding delay measures the time steps that one has to wait for the message to be decoded at the receiver with a small enough probability of decoding error. The short coding delay is also reflected in Fig. 9 (c), where we compare the message and the decoded message bit by bit and count how many bits are correctly obtained by the receiver. At time $k = 24$, the channel can transmit 35.8 bits if at each step the capacity C is attained, and the simulation shows that on average 34.9 bits are actually correctly decoded.
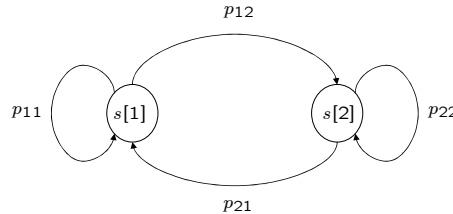

Fig. 8.   Gilbert-Elliot fading channel with AWGN.

Therefore, though the availability of output feedback at the transmitter does not affect the capacity when DTRCSI is available, we have seen that, output feedback can considerably simplify the coding

---

[12]Note again that PE_k, the probability of error averaged over Ω_k, decays only singly exponentially. [45] presented a coding scheme which can reduce the decoding errors caused by atypical channel state sequences for "streaming" communication. Whether the same idea is applicable here to lead to a doubly exponential decay of the averaged error probability remains to be seen.
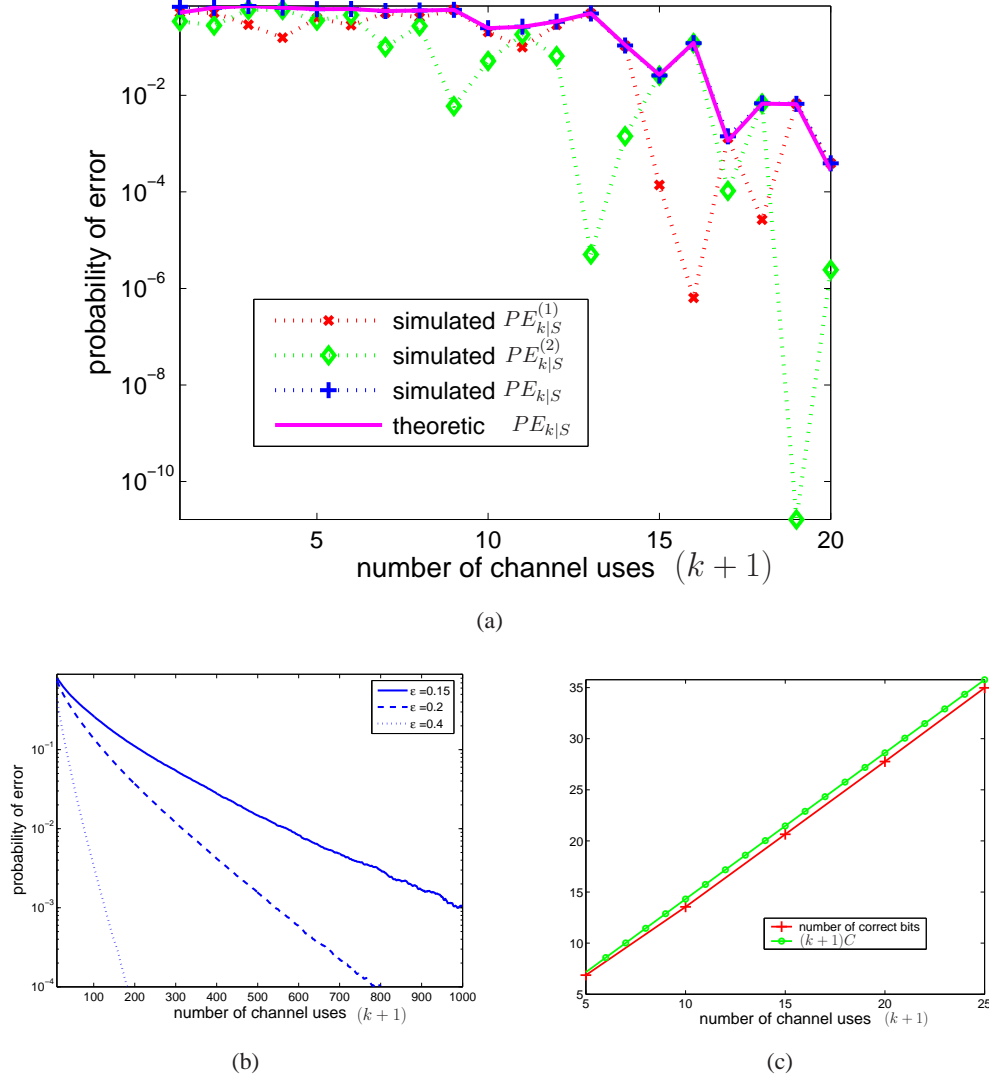
Fig. 9. (a) Simulated $PE_{k|S}^{(j)}$, simulated $PE_{k|S}$, and theoretic $PE_{k|S}$. (b) Theoretic PE_k. (c) The number of bits that has been correctly decided and the number of bits that could be correctly decided if at each step the capacity rate is attained. $s[1] = 2$, $s[2] = 1$, $p_{11} = 0.65$, $p_{22} = 0.38$, $P = 3$, and $\epsilon = 0.2$ (i.e. $R = 0.8C$), unless otherwise specified in the legend.

design and coding process while achieving a rate arbitrarily close to the capacity (in contrast to some sophisticated designs to approach the capacity using Turbo codes or LDPC codes, see e.g. [46]), and it leads to a better performance in terms of probability of error than the coding schemes without feedback in the literature. Recall that turbo codes or LDPC codes typically need long coding lengths of at least several thousands to achieve a reasonable performance (in [46], coding length of 8000 was used over a Markov channel). However, since power adaptation has not been employed in those schemes in the existing literature, a fair and more accurate comparison with our scheme is not yet available.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed, based on a control-oriented approach, a capacity-achieving feedback communication scheme for an AFSMC with precise CSI available to the receiver immediately and to the transmitter with unit delay. In essence, this scheme consists of a set of subsystems designed for an AWGN channel with feedback, and the subsystems are multiplexed to share the AFSMC and the feedback link according to the augmented channel states. The scheme greatly simplifies the complexity

in the coding design and coding processes. The error probability decreases to zero doubly exponentially for typical channel state sequences and it shortens the coding length, compared with existing coding schemes over an AFSMC in the literature.

In constructing the proposed scheme, we established the equivalence between feedback stabilization over an AFSMC and communication with access to noiseless feedback over the same channel. We have seen that the utilization of the control-theoretic equivalence of the proposed coding scheme facilitates the development. In fact, the ideas of introducing multiplexing and augmented channel states in the feedback communication problem originated from the study of the associated control problem; particularly the idea of multiplexing was motivated by the control of MJLS, and the idea of augmenting channel states was motivated by considering control systems with delay. We remark that these two ingredients may play a significant role in studying more general feedback communication systems with known channel variation (varying with time) and delays.

There are several open directions for future work. First, note that the assumption of perfect (CSI and output) feedback is a major limitation in existing feedback communication studies. The literature on imperfect CSI at the transmitter side is vast; see e.g. [3], [8], [9], [47]–[50]. Quantized CSI feedback is typically assumed, and usually waterfilling-type capacity formulas are obtained, in which case one may extend the multiplexing designs to achieve the capacities. If, on the other hand, the transmitter-side CSI is corrupted by sources other than quantization, it remains to be studied how the multiplexing used in this paper may be extended, as the receiver may not know how the sub-codewords are multiplexed at the transmitter side. However, the fundamental difficulty in noisy feedback lies in the imperfect output feedback, as shown in [10], [11]. In the noisy output feedback case, the SK coding scheme (as well as its direct extensions) leads to a non-vanishing probability of error for a fixed signalling rate, or leads to a vanishing signalling rate if the probability of error decays to zero, as the coding length tends to infinity. Nevertheless, it would be useful to study the more realistic problem of communication with noisy feedback as the ideal problem of communication with noiseless feedback has been resolved.

In addition, we wish to further explore the role of the cheap control (or its counterpart in estimation theory, the Kalman filter) in feedback communication. The Kalman filter, the optimal recursive MMSE estimator, can be easily transformed into an optimal feedback communication system achieving the Shannon capacity, for either an AWGN channel with output feedback or an AFSMC with output feedback and side information (cf. [27]). Moreover, our existing study has revealed further tight connections among communication, estimation, and control, which may be seen as an extension of the connections between communication and control discovered in e.g. [12] and have been employed to characterize the capacity-achieving feedback schemes for Gaussian channels with memory (cf. [16], [51]).

Finally, we wish to extend the main ideas of this paper to MIMO time-varying fading channels with noiseless output feedback. For the case of single-user MIMO Gaussian channels with perfect CSI available to the receiver instantaneously and to the transmitter with delay, we expect that our proposed communication scheme can be extended relatively easily. When multiple users are present, however, more complicated strategies may be needed. The challenges may be seen from the fact that only bounds for the capacities of some multiple-user channels are available. Multiple-user MIMO problems are especially challenging. Not surprisingly, challenges present in both the communication problem and the associated control problems, especially for some multiple-user MIMO systems. From a control theoretic perspective, [12] obtained the feedback capacity or best-known feedback capacity regions for some multiple-user time-invariant MIMO Gaussian channels, including multiple-access channels, broadcast channels, and symmetric interference channels, and feedback communication schemes to achieve the capacity or capacity bounds were also given. However, open problems in feedback communication translate to open problems in feedback control, For example, the capacity problem of Gaussian interference channel with feedback becomes a structured control optimization problem ([12]). At any rate, applying the

combination of the techniques developed in this paper for SISO time-varying channels and those developed in [12] for multiple-user MIMO time-invariant channels may serve as an initial step to tackle the MIMO time-varying feedback communication problems, and may provide useful insights, possibly on both feedback information theory and control theory.

APPENDIX

A. Proof of $C = \log \tilde{a}$

Assume without loss of generality that $S_k = s[i]$ and $S_{k+1} = s[j]$, both drawn from the stationary distribution of $\{S_k\}$. Then it holds that

$$
\begin{aligned}
C &= \frac{1}{2}\mathbf{E}_{S_k,S_{k+1}} \log\left(1 + (S_{k+1})^2\gamma(S_k)\right) \\
&= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\pi[i]p_{ij}\log\left(1 + s[j]^2\gamma(s[i])\right) \\
&= \sum_{i=1}^{m}\left(\sum_{j=1}^{m}\pi[i]p_{ij}\log a(s[i],s[j])\right) \\
&= \sum_{i=1}^{m}\log\bar{a}(s[i]) \\
&= \log\tilde{a}
\end{aligned}
$$

B. Proof of converse

This proof is motivated by the converse proofs in [4] and in [8].

For any $(M, K+1)$ code, the Fano's inequality yields that

$$
\begin{aligned}
h(W|y_0^K, S_0^K) &= h(W) - I(W; y_0^K, S_0^K) \\
&= \log M - I(W; y_0^K, S_0^K) \\
&\leq h(PE_K) + PE_K \log M
\end{aligned}
\tag{88}
$$

and hence that

$$
\frac{1}{K+1}\log M \leq \frac{1}{(K+1)(1-PE_K)}\left(h(PE_K) + I(W; y_0^K, S_0^K)\right),
\tag{89}
$$

where W is the uniformly distributed message. If the code leads to a vanishing probability of error as K tends to infinity, namely PE_K (and hence $h(PE_K)$) is vanishing, then the signalling rate R satisfies

$$
R \leq \limsup_{K\to\infty}\frac{1}{K+1}I(W; y_0^K, S_0^K).
\tag{90}
$$

In addition, the code needs to satisfy the power constraint

$$
\limsup_{K\to\infty}P_K = \limsup_{K\to\infty}\frac{1}{K+1}\mathbf{E}\sum_{k=0}^{K}(u_k)^2 \leq \mathcal{P}.
\tag{91}
$$

To see that the above equality is true, by the definition of P_K, we have

$$
P_K = \frac{1}{K+1}\sum_{k=0}^{K}\left(\sum_{S_0^{K-1}\in\Omega_{K-1}}\left(\Pr(S_0^{K-1})\mathbf{E}(u_{k|S})^2\right)\right).
\tag{92}
$$

30

Note that $\mathbf{E}(u_{k|S})^2$ depends only on S_0^{k-1}, by a simple counting argument, it holds that

$$
\begin{aligned}
P_K &= \frac{1}{K+1}\sum_{k=0}^{K}\left(\sum_{S_0^{k-1}\in\Omega_{k-1}}\left(\Pr(S_0^{k-1})\mathbf{E}(u_{k|S})^2\right)\right) \\
&= \frac{1}{K+1}\sum_{k=0}^{K}\mathbf{E}(u_k)^2.
\end{aligned}
\tag{93}
$$

Thus (91) holds.

Now note that u_k depends on S_0^{k-1} and y_0^{k-1}. We then denote $\mathbf{E}u_k^2$, conditioned on S_0^{k-1} and y_0^{k-1}, as $f_{k|S_0^{k-1},y_0^{k-1}}$. Then $\sum_{k=0}^{K}\mathbf{E}f_{k|S_0^{k-1},y_0^{k-1}}=(K+1)P_K$ in light of (48). Since

$$
\begin{aligned}
&I(W;y_0^K,S_0^K)\\
=\;&I(W;y_0^K|S_0^K)+I(W;S_0^K)\\
\overset{(a)}{=}\;&I(W;y_0^K|S_0^K)\\
=\;&h(y_0^K|S_0^K)-h(y_0^K|S_0^K,W)\\
=\;&\sum_{k=0}^{K}\left(h(y_k|S_0^{k-1},y_0^{k-1},S_k^K)-h(y_k|S_0^{k-1},y_0^{k-1},W,S_k,S_{k+1}^K)\right)\\
=\;&\sum_{k=1}^{K}\left(h(y_k|S_0^{k-1},y_0^{k-1},S_k^K)-h(y_k|S_0^{k-1},y_0^{k-1},W,S_k,S_{k+1}^K)\right)+h(y_0|S_0)-h(y_0|S_0,W)\\
\overset{(b)}{\leq}\;&\sum_{k=1}^{K}\left(h(y_k|S_{k-1},y_0^{k-1},S_k)-h(N_k)\right)+I(W;y_0|S_0)\\
\overset{(c)}{=}\;&\sum_{k=1}^{K}\left(h(y_k|S_{k-1},y_0^{k-1},S_k)-h(y_k|S_{k-1},y_0^{k-1},S_k,u_k)\right)+I(W;y_0|S_0)\\
\overset{(d)}{\leq}\;&\sum_{k=1}^{K}I(u_k;y_k|S_{k-1},y_0^{k-1},S_k)+I(u_0;y_0|S_0)\\
\overset{(e)}{\leq}\;&\frac{1}{2}\sum_{k=1}^{K}\mathbf{E}\log\left(1+(S_k)^2 f_{k|S_0^{k-1},y_0^{k-1}}\right)+\frac{1}{2}\mathbf{E}\log\left(1+(S_0)^2 f_0\right)\\
=\;&\frac{1}{2}\sum_{k=1}^{K}\mathbf{E}\left\{\mathbf{E}\left[\log\left(1+(S_k)^2 f_{k|S_0^{k-1},y_0^{k-1}}\right)\Big|S_{k-1},S_k\right]\right\}+\frac{1}{2}\mathbf{E}\log\left(1+(S_0)^2 f_0\right)\\
\overset{(f)}{\leq}\;&\frac{1}{2}\sum_{k=1}^{K}\mathbf{E}\log\left[1+(S_k)^2\mathbf{E}\left(f_{k|S_0^{k-1},y_0^{k-1}}\Big|S_{k-1}\right)\right]+\frac{1}{2}\mathbf{E}\log\left(1+(S_0)^2 f_0\right)\\
:=\;&\frac{1}{2}\sum_{k=1}^{K}\mathbf{E}\log\left(1+(S_k)^2\Gamma(S_{k-1})\right)+\frac{1}{2}\mathbf{E}\log\left(1+(S_0)^2 f_0\right)
\end{aligned}
\tag{94}
$$

where (a) follows from the independence between W and S_0^K, (b) is due to that conditioning reduces entropy and the definitions of u_k and y_k, (c) is again due to the definitions of u_k and y_k, (d) is because of the data processing inequality, (e) is because Gaussian input maximizes mutual information under power constraint and conditioned on channel state and past channel outputs, and (f) follows from Jensen's inequality, the independence between S_k and y_0^{k-1} conditioned on S_{k-1}, and the independence between S_k and S_0^{k-2} conditioned on S_{k-1}.

Therefore, we have that

$$
I(W;y_0^K,S_0^K)\leq\frac{1}{2}\sum_{k=1}^{K}\mathbf{E}\log\left(1+(S_k)^2\Gamma(S_{k-1})\right)+\frac{1}{2}\mathbf{E}\log(1+(S_0)^2 f_0)
\tag{95}
$$

subject to power constraint $\sum_{k=0}^{K} \mathbf{E}\Gamma(S_{k-1}) \le (K+1)P_K$. Hence, by the stationarity and ergodicity of the channel state process, it holds that

$$R \le \frac{1}{2}\mathbf{E}\log\left(1 + (S_k)^2\Gamma(S_{k-1})\right) \tag{96}$$

where S_{k-1} follows the stationary distribution and $\mathbf{E}\Gamma(S_{k-1}) \le \mathcal{P}$. Finally we have $R \le C$ by the optimality of $\gamma(\cdot)$.

C. Proof of Lemma 2

i) The dynamics of $x_k^{(j)}$ without external input is $x_{k+1}^{(j)} = a(S_{k-1}, S_k)^{-1}x_k^{(j)}$ if $S_{k-1} = s[j]$, or $x_{k+1}^{(j)} = x_k^{(j)}$ otherwise. So we have

$$x_{k+1}^{(j)} = \phi_k^{(j)}x_0^{(j)}$$

and hence Φ_k is the state transition matrix for (23). To show the convergence in probability, we need to show that for any $\epsilon > 0$ small enough, it holds that

$$\Pr\left(|\phi_k^{(j)}| < \epsilon\right) \to 1, \tag{97}$$

in other words,

$$\Pr\left(\left|\prod_{l=1}^{m} a(s[j], s[l])^{-n(j,l,k)}\right| < \epsilon\right) \to 1. \tag{98}$$

Since $|a(s[j], s[l])| > 1$ for all j and l, by the Continuous Mapping Theorem, it is sufficient to show that for any $M > 0$ large enough, there exists at least one l such that as k goes to infinity,

$$\Pr\left(n(j, l, k) > M\right) \to 1. \tag{99}$$

Notice that $p_{jl} > 0$ for some l, by (6), (99) is true because

$$\Pr\left(n(j, l, k) > (k+1)(\pi[j]p_{jl} - \delta)\right) \to 1 \tag{100}$$

for any $\delta > 0$ sufficiently small (noting that $\pi[j] > 0$ by ergodicity).

ii) Conditioned on S_0^k and $\boldsymbol{x}_0$, we obtain that

$$\mathbf{E}x_{k+1}^{(j)} = \phi_k^{(j)}x_0^{(j)}, \tag{101}$$

and hence, by i), the boundedness of $\mathbf{E}\boldsymbol{x}_k^{(j)}$ on Ω for each k as well as the convergence in probability of $\mathbf{E}x_k^{(j)}$ to zero.

iii) We first show the independence of $x_k^{(j)}$ and $x_k^{(l)}$ when $j \neq l$. To show this, notice that, conditioned on $\boldsymbol{x}_0$ and $\{S_k\}$, both $x_k^{(j)}$ and $x_k^{(l)}$ are Gaussian; this is because that the system is linear and that the only randomness in $x_k^{(j)}$ and $x_k^{(l)}$ is from the AWGN $\{N_k\}$. Then it is only needed to show they are uncorrelated, namely

$$\mathbf{E}x_k^{(j)}x_k^{(l)} = \mathbf{E}x_k^{(j)}\mathbf{E}x_k^{(l)}. \tag{102}$$

If $k = 0$, obviously (102) holds. Suppose that (102) holds for some k, then for $(k+1)$ and for any i,

$$x_{k+1}^{(i)} = \begin{cases} a(S_{k-1}, S_k)^{-1}x_k^{(i)} - b(S_{k-1}, S_k)N_k & \text{if } S_{k-1} = s[i] \\ x_k^{(i)} & \text{otherwise.} \end{cases} \tag{103}$$

So

$$
\mathbf{E}x_{k+1}^{(j)}x_{k+1}^{(l)} = \begin{cases} \mathbf{E}x_k^{(j)}x_k^{(l)} & \text{if } S_{k-1} \neq s[j] \text{ and } S_{k-1} \neq s[l] \\ \mathbf{E}[a(S_{k-1}, S_k)^{-1}x_k^{(j)} - b(S_{k-1}, S_k)N_k]x_k^{(l)} & \text{if } S_{k-1} = s[j] \\ \mathbf{E}x_k^{(j)}[a(S_{k-1}, S_k)^{-1}x_k^{(l)} - b(S_{k-1}, S_k)N_k] & \text{if } S_{k-1} = s[l] \end{cases}
$$

$$
= \begin{cases} \mathbf{E}x_{k+1}^{(j)}\mathbf{E}x_{k+1}^{(l)} & \text{if } S_{k-1} \neq s[j] \text{ and } S_{k-1} \neq s[l] \\ \mathbf{E}a(S_{k-1}, S_k)^{-1}x_k^{(j)}x_k^{(l)} & \text{if } S_{k-1} = s[j] \\ \mathbf{E}x_k^{(j)}a(S_{k-1}, S_k)^{-1}x_k^{(l)} & \text{if } S_{k-1} = s[l] \end{cases}
$$

$$
= \begin{cases} \mathbf{E}x_{k+1}^{(j)}\mathbf{E}x_{k+1}^{(l)} & \text{if } S_{k-1} \neq s[j] \text{ and } S_{k-1} \neq s[l] \\ \mathbf{E}a(S_{k-1}, S_k)^{-1}x_k^{(j)}\mathbf{E}x_k^{(l)} & \text{if } S_{k-1} = s[j] \\ \mathbf{E}x_k^{(j)}\mathbf{E}a(S_{k-1}, S_k)^{-1}x_k^{(l)} & \text{if } S_{k-1} = s[l] \end{cases}
$$

$$
= \mathbf{E}x_{k+1}^{(j)}\mathbf{E}x_{k+1}^{(l)}. \tag{104}
$$

Thus, (102) holds for any $k \in \mathbb{N}$ by induction.

By the independence of $x_k^{(j)}$ and $x_k^{(l)}$ when $j \neq l$, Σ_k is a diagonal matrix. If $S_{k-1} = s[j]$ then

$$
\begin{aligned}
\mathbf{E}(x_{k+1}^{(j)})^2 &= a(S_{k-1}, S_k)^{-2}\mathbf{E}(x_k^{(j)})^2 + b(S_{k-1}, S_k)^2 \\
&\leq a^{-2}\mathbf{E}(x_k^{(j)})^2 + b^2,
\end{aligned}
$$

where $a := \min_{j,l} a(s[j], s[l])$ and $b := \max_{j,l} |b(s[j], s[l])|$; or if, however, $S_{k-1} \neq s[j]$, then

$$
\mathbf{E}(x_{k+1}^{(j)})^2 = \mathbf{E}(x_k^{(j)})^2.
$$

Since $0 < a^{-1} < 1$, for any k,

$$
\begin{aligned}
\mathbf{E}(x_k^{(j)})^2 &\leq a^{-2K}(x_0^{(j)})^2 + \frac{1 - a^{-2K}}{1 - a^{-2}}b^2 \\
&\leq (x_0^{(j)})^2 + \frac{1}{1 - a^{-2}}b^2 \\
&\leq 0.5^2 + \frac{1}{1 - a^{-2}}b^2,
\end{aligned} \tag{105}
$$

where $K \geq 0$ (dependent on k and $\{S_k\}$). Since $\mathbf{E}(x_k^{(j)})^2 \geq \underline{b}^2$ where $\underline{b} := \min_{j,l} |b(s[j], [l])| > 0$, $\mathbf{E}(x_k^{(j)})^2$ is bounded from both above and below by some positive constants for all k; note that the constants can be chosen as independent on k, $\boldsymbol{x}_0$, and $\{S_k\}$. Notice that $\Sigma_k^{(j)} := \mathbf{E}(x_k^{(j)})^2 - (\mathbf{E}x_k^{(j)})^2$ is strictly positive since the randomness in the noise enters the system. Then, because $|\mathbf{E}x_k^{(j)}|$ decreases to zero monotonically as k increases, $\Sigma_k^{(j)}$ is uniformly bounded from above and from below by positive constants for any $\boldsymbol{x}_0$, $\{S_k\}$, and k.

D. Proof of Proposition 2

To prove this proposition, first note (91) holds. It is sufficient to show that

$$
\mathbf{E}(u_k)^2 \to \sum_{j=1}^{m} \pi[j]\gamma(s[j]). \tag{106}
$$

Then by the Cesaro mean (i.e., if a_n converges to a, then the average of the first n terms converges to a as n goes to infinity), the limit in the right-hand side of (91) exists and the result follows.

Now let us study the recursion for $\mathbf{E}(x_k^{(j)})^2$. Assume $S_{k-2} = s[j]$ and $S_{k-1} = s[l]$. Then (26) yields that

$$
\begin{aligned}
\mathbf{E}(x_k^{(j)})^2 &= a(s[j], s[l])^{-2} \mathbf{E}(x_{k-1}^{(j)})^2 + b(s[j], s[l])^2 \\
&\overset{(a)}{=} a(s[j], s[l])^{-2} \mathbf{E}(x_{k-1}^{(j)})^2 + a(s[j], s[l])^{-2} \gamma(s[j])^2 s[l]^2 \\
&= a(s[j], s[l])^{-2} (\mathbf{E}(x_{k-1}^{(j)})^2 + \gamma(s[j])^2 s[l]^2),
\end{aligned}
$$

where (a) follows from (21). Subtracting both sides by $\gamma(s[j])$, we obtain that

$$
\begin{aligned}
\mathbf{E}(x_k^{(j)})^2 - \gamma(s[j]) &= a(s[j], s[l])^{-2} \left(\mathbf{E}(x_{k-1}^{(j)})^2 + \gamma(s[j])^2 s[l]^2 - a(s[j], s[l])^2 \gamma(s[j]) \right) \\
&= a(s[j], s[l])^{-2} \left(\mathbf{E}(x_{k-1}^{(j)})^2 - \gamma(s[j]) \right),
\end{aligned}
$$

and thus conditioned on a channel state sequence S_0^k,

$$
\mathbf{E}(x_k^{(j)})^2 - \gamma(s[j]) = (\phi_{k-1}^{(j)})^2 \left(\mathbf{E}(x_0^{(j)})^2 - \gamma(s[j]) \right). \tag{107}
$$

It follows from Lemma 2 i) that $(\mathbf{E}(x_k^{(j)})^2 - \gamma(s[j]))$ converges in probability to zero, namely $\mathbf{E}(x_k^{(j)})^2$ converges to $\gamma(s[j])$ in probability. Note that we have omitted the conditioning on S_0^k of $x_k^{(j)}$ to simplify notation; the obtained $\mathbf{E}(x_k^{(j)})^2$ is a random variable depending on the random variable S_0^k.

Since the channel input u is the multiplexing of $x_k^{(j)}$, the power of u is the power of $x_k^{(j)}$ averaged over j, therefore we have that

$$
\mathbf{E}(u_k)^2 \overset{\mathrm{P}}{\to} \sum_{j=1}^{m} \pi[j] \gamma(s[j]), \tag{108}
$$

which implies (106) as $\mathbf{E}(u_k)^2$ is deterministic. Thus we complete this proof.

E. Proof of Proposition 3

We first strengthen Lemma 2.

Lemma 3. *Assume the hypotheses of Theorem 1, and fix a channel state sequence $\{S_k\}$ in Ω. Then for the control setup (23),*
  i) It holds that for each j,

$$
\lim_{k \to \infty} \phi_k^{(j)} \to 0 \text{ if } \{S_k\} \in \Omega_{TYP}; \tag{109}
$$

  ii) For any fixed initial condition $\boldsymbol{x}_0$, it holds that

$$
\lim_{k \to \infty} \mathbf{E} \boldsymbol{x}_k \to 0 \text{ if } \{S_k\} \in \Omega_{TYP}. \tag{110}
$$

The proof of the lemma is as follows. By (77), on Ω_{TYP}, $n(j, k)$ goes to infinity as k goes to infinity, since $\pi[j] > 0$. Because $0 < a(S_{k-1}, S_k)^{-1} < 1$, $\phi_k^{(j)}$ goes to zero on Ω_{TYP}. Moreover, conditioned on S_0^k and $\boldsymbol{x}_0$, we obtain that

$$
\mathbf{E} x_{k+1}^{(j)} = \phi_k^{(j)} x_0^{(j)} \tag{111}
$$

and hence $\mathbf{E} \boldsymbol{x}_k^{(j)}$ converges to zero on Ω_{TYP}. $\qquad \square$

Now we are ready to prove the proposition. Fix any $\{S_k\} \in \Omega_{TYP}$. Recall that

$$
PE_{k|S}^{(j)} = Q_{k,1} + Q_{k,2}, \tag{112}
$$

where $Q_{k,1}$ satisfies

$$
Q_{k,1} \le Q \left(\frac{\xi_k^{(j)}}{2\sqrt{\Sigma_{k+1}^{(j)}}} \alpha^{n(j,k) + (k+1)\left(\frac{1}{\epsilon} - 1\right)\left(\pi[j] - \frac{n(j,k)}{k+1}\right)} + \frac{x_0^{(j)} \phi_k^{(j)}}{\sqrt{\Sigma_{k+1}^{(j)}}} \right), \tag{113}
$$

34

and $Q_{k,2}$ satisfies a similar inequality, for sufficiently large k. This is because for any $\mu > 0$, there exists k large enough such that the prefix S_0^k of $\{S_k\}$ is in $\Omega_{k,\mu}$, then the result in the proof of Proposition 1 can be used here. Let

$$\delta_k := (k+1)\left(\frac{1}{\epsilon}-1\right)\left(\pi[j]-\frac{n(j,k)}{k+1}\right)+\log_\alpha\left(\frac{\xi_k^{(j)}+2\alpha^{-n(j,k)\left[1+\left(\frac{1}{\epsilon}-1\right)\left(\frac{(k+1)\pi[j]}{n(j,k)}-1\right)\right]}x_0^{(j)}\phi_k^{(j)}}{2\sqrt{\Sigma_{k+1}^{(j)}}}\right).$$
(114)

Then

$$Q_{k,1} = Q\left(\alpha^{n(j,k)+\delta_k}\right),$$
(115)

Since $n(j,k)/(k+1)$ converges to $\pi[j]$ when k goes to infinity for any $\{S_k\} \in \Omega_{TYP}$, we have that $\delta_k/(k+1)$ vanishes when k goes to infinity (noting that δ_k is not viewed as a random variable here since $\{S_k\}$ is viewed as realization). That is,

$$Q_{k,1} \le Q\left(\alpha^{n(j,k)+o(k+1)}\right).$$
(116)

The Chernoff bound of the Q-function (cf. [52]) says that

$$Q(t) \le \frac{1}{\sqrt{2\pi}t}\exp(-\frac{1}{2}t^2) = \exp\left(-\frac{1}{2}\left(t^2+\log(2\pi t^2)\right)\right).$$

Then

$$\begin{aligned}
Q_{k,1} &\le \exp\left\{-\frac{1}{2}\left[\alpha^{2n(j,k)+o(k+1)}+\log(2\pi\alpha^{2n(j,k)+o(k+1)})\right]\right\}\\
&= \exp\left\{-\frac{1}{2}(\alpha^2)^{n(j,k)+\frac{1}{2}\log_\alpha\left[\alpha^{o(k+1)}+\alpha^{-2n(j,k)}\log(2\pi\alpha^{2n(j,k)+o(k+1)})\right]}\right\}\\
&= \exp\left(-\frac{1}{2}(\alpha^2)^{\pi[j](k+1)+o(k+1)}\right).
\end{aligned}$$

Similarly,

$$Q_{k,2} := Q\left(\frac{0.5\xi_k^{(j)}}{\bar{a}[j]^{(k+1)(1-\epsilon)}\phi_k^{(j)}\sqrt{\Sigma_{k+1}^{(j)}}}-\frac{x_0^{(j)}\phi_k^{(j)}}{\sqrt{\Sigma_{k+1}^{(j)}}}\right) \le \exp\left(-\frac{1}{2}(\alpha^2)^{\pi[j](k+1)+o(k+1)}\right),$$

and hence,

$$PE_{k|S}^{(j)} \le 2\exp\left(-\frac{1}{2}(\alpha^{2\pi[j]})^{(k+1)+o(k+1)}\right).$$
(117)

To get the (asymptotic) decay exponent of $PE_{k|S}^{(j)}$, noticing that for k large enough, the Chernoff bound becomes tight, we can derive (following the steps similar to above) that

$$\begin{aligned}
\lim_{k\to\infty}\frac{1}{k+1}\log\left(\log\left(\frac{1}{Q_{1,k}}\right)\right) &= \lim_{k\to\infty}-2\sum_{l=1}^{m}d(j,l,k)\log(a(s[j],s[l]))\\
&= 2\epsilon\sum_{l=1}^{m}\pi[j]p_{jl}\log(a(s[j],s[l]))\\
&= 2\epsilon\log\bar{a}[j].
\end{aligned}$$

It can be also shown that

$$\lim_{k\to\infty}\frac{1}{k+1}\log\left(\log\left(\frac{1}{Q_{2,k}}\right)\right) = 2\epsilon\log\bar{a}[j].$$

It is easily seen that

$$\lim_{k\to\infty}\frac{1}{k}\left(\log\left(\frac{1}{\exp(-a^k)+\exp(-b^k)}\right)\right) = \min(\log a, \log b)$$
(118)

35

for $a, b > 0$. (That is, the "average" decay exponent of $(\exp(-a^k) + \exp(-b^k))$ is the smaller decay exponent for $\exp(-a^k)$ and $\exp(-b^k)$.) Then (80) follows from (118). Finally, notice that asymptotically $PE_{k|S} = \sum_{j=1}^{m} PE_{k|S}^{(j)}$, then (81) follows from (118).

F. Modifications if A1 or A2 does not hold

1) When A1 is not assumed: In this case, there exists some states that are assigned with zero power. Suppose $s[J]$ is the only such state; for cases with more than one such states, the same idea applies. We now have $a(s[i], s[J]) = 1$ and $\bar{a}[J] = 1$. Then the Jth subsystem needs some modifications, as described below. For the encoding, let $x_0^{(J)} = 0$ and make it known to the receiver. That is, the Jth subsystem is not used to transmit any message. It leads to zero signalling rate, zero transmission power, and zero probability of error associated with this subsystem. We then see that in this situation, our scheme behaves exactly as the capacity solution equation suggests. Namely, at the moment when no power is assigned, no message needs to be transmitted, no probability of error is incurred, and there is no contribution to the rate. Therefore, it can be easily verified that any rate below the capacity is achievable by simply noting that: 1) In Lemma 2, i) and ii) hold for any $j \neq J$, and iii) holds if we modify the definitions of a and b as $a := \min_{j \neq J, l} a(s[j], s[l])$ and $b := \max_{j \neq J, l} |b(s[j], s[l])|$; 2) $C = \sum_{i=1}^{m} \log \bar{a}(s[i]) = \sum_{i \neq J} \log \bar{a}(s[i])$; 3) $\mathbf{E}u^2 = \sum_{i=1}^{m} \pi[i]\mathbf{E}(x^{(i)})^2 = \sum_{i \neq J} \pi[i]\mathbf{E}(x^{(i)})^2$; and 4) Proposition 1 holds if we assign $PE_{k|S}^{(j)} = 0$.

2) When A2 is not assumed: In this case there is some i with $s[i] = 0$. Then $a(s[j], s[i]) = 1$, but $\bar{a}[j] > 1$ still holds for each j with $\gamma(s[j]) \neq 0$. To see this, assume otherwise for some J, $\bar{a}[J] = 1$ but $\gamma(s[J]) > 0$. By (29) and (20), this would yield that, for each $l = 1, \cdots, m$, it must hold either i) $s[l] = 0$ or ii) $\pi[J]p_{Jl} = 0$. However, ii) is equivalent to $p_{Jl} = 0$ since $\pi[J] > 0$. Hence, for any l such that $s[l] \neq 0$, p_{Jl} has to be 0; i.e., the probability of $s[J]$ jumping to any nonzero state is *zero*. In other words, $s[J]$ must jump to $s[J]$ with probability one, which according to [4] would imply that $\gamma(s[J]) = 0$, a contradiction. So $\bar{a}[j] > 1$ holds as long as $\gamma(s[j]) \neq 0$, and $a(s[j], s[l]) > 1$ and $\pi[j]p_{jl} > 0$ for some l. Thus, $\phi_k^{(j)}$ converges to zero following the proof used in Lemma 2 i). Again iii) holds if we modify the definitions of a and b as $a := \min_{j \neq J, l} a(s[j], s[l])$ and $b := \max_{j \neq J, l} |b(s[j], s[l])|$. For the proof of Proposition 1, modify the definition of a as $a := \min_{l \neq i} a(s[j], s[l]) > 1$. Then Lemma 2 as well as the main results hold.

G. Proof for the case of AWGN i.i.d. fading with infinite state

Pick any $\epsilon > 0$ small enough. Uniformly partition the unit interval $[-\frac{1}{2}, \frac{1}{2}]$ into $\lfloor M_k \rfloor$ sub-intervals, where

$$M_k := \exp((k+1)(1-\epsilon)\mathbf{E}\log A). \tag{119}$$

Then the asymptotic signalling rate is

$$
\begin{aligned}
R &= \lim_{k \to \infty} \left(\frac{\log M_k}{k+1} - \frac{\log \xi_k}{k+1} \right) \\
&= \lim_{k \to \infty} \frac{(k+1)(1-\epsilon)\mathbf{E}\log A}{k+1} \\
&= (1-\epsilon)\mathbf{E}\log A \\
&= (1-\epsilon)\mathbf{E}\frac{1}{2}\log(1 + S^2\mathcal{P}) \\
&= (1-\epsilon)C_{iid,inf},
\end{aligned}
\tag{120}
$$

where $\xi_k := M_k / \lfloor M_k \rfloor$.

In the i.i.d. case with infinite state, a channel state will be re-visited with zero probability, and hence the idea used for AFSMC cannot be applied; notations such as $n(j, l, k)$ and $\phi_k^{(j)}$ that were used to show stability and vanishing probability of error for AFSMC make no sense here. The infinite-state i.i.d. case, however, requires only a simpler notion of "typicality" which is still based on the Weak Law of Large Numbers to establish the achievable rate result. Define the (k, μ)-*typical set* of sequences S_0^k as

$$\Omega_{k,\mu} := \left\{ S_0^k \left\| \frac{1}{k+1} \sum_{t=0}^{k} \log A(S_t) - \mathbf{E} \log A \right| < \mu \right\}. \tag{121}$$

As S_t forms an i.i.d. process, so does $\log A(S_t)$ and thus, by the Weak Law of Large Numbers, it holds that $\Pr(\Omega_{k,\mu}) \to 1$.

Following the idea in the finite state-space case, to show the reliable communication with rate R, it suffices to show that the associated control system is stabilized in the sense of bounded and vanishing first moment $\mathbf{E}x_k$ and bounded second moment $\mathbf{E}(x_k - \mathbf{E}x_k)^2$, for any typical channel state sequence. These stability notions of the control system imply that the receiver estimate, $\hat{x}_{0,k}$, is "close" to the transmitted codeword, x_0; recall that both $\hat{x}_{0,k}$ and x_0 are scalars. Then we need to show that this "closeness" can be translated into vanishing probability of decoding error and thus the achievability of the signalling rate R.

Now we show that the stability of the control system can be easily proven; for brevity we only show that $\mathbf{E}x_k$ is bounded and vanishing here. Note that

$$x_k = A(S_{k-1})^{-1}x_{k-1} - b(S_{k-1})N_{k-1}. \tag{122}$$

The first moment evolves according to

$$\mathbf{E}x_k = A(S_{k-1})^{-1}\mathbf{E}x_{k-1} \tag{123}$$

when conditioned on any typical channel state sequence and initial condition, and hence

$$\mathbf{E}x_k = \prod_{t=0}^{k} A(S_{t-1})^{-1}x_0. \tag{124}$$

Then the convergence of the first moment follows from (121) and that $\mathbf{E} \log A > 0$. To show that the stability translates to vanishing probability of error, we follow the steps in the Markov case and note that we need to show

$$\exp\left[(k+1)\left((1-\epsilon)\mathbf{E}\log A - \frac{1}{k+1}\sum_{t=0}^{k}\log A(S_t)\right)\right] \tag{125}$$

is vanishing, which is true for sufficiently small $\mu > 0$ in view of (121) and the ϵ slack used. Finally, the power computation can be done as before. Thus it follows that the proposed scheme is optimal.

REFERENCES

[1] A. J. Goldsmith and P. P. Varaiya. Capacity, mutual information, and coding for finite-state Markov channels. *IEEE Trans. Inform. Theory*, 42(3):868–886, May 1996.

[2] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Inform. Theory*, 43(6):1986–1992, Nov. 1997.

[3] G. Caire and S. Shamai (Shitz). On the capacity of some channels with channel state information. *IEEE Trans. Inform. Theory*, 45(6):2007–2019, Sept. 1999.

[4] H. Viswanathan. Capacity of Markov channels with receiver CSI and delayed feedback. *IEEE Trans. Inform. Theory*, 45(2):761–771, March 1999.

[5] R. Negi and J. M. Cioffi. Delay-constrained capacity with causal feedback. *IEEE Trans. Inform. Theory*, 48(9):2478–2494, Sept. 2002.

[6] S. Yang, A. Kavcic, and S. Tatikonda. Feedback capacity of finite-state machine channels. *IEEE Trans. Inform. Theory*, 51(3):799–810, Mar. 2005.

[7] T. Holiday, A. J. Goldsmith, and P.Glynn. Capacity of finite state channels based on lyapunov exponents of random matrices. *IEEE Trans. Inform. Theory*, 52(8):3509 – 3532, Aug. 2006.

[8] M. A. Sadrabadi, M. A. Maddah-Ali, and A. K. Khandani. On the capacity of time-varying channels with periodic feedback. *IEEE Trans. Inform. Theory*, 53(8):2910–2915, Aug. 2007.

[9] A. J. Goldsmith and M. Medard. Capacity of time-varying channels with causal channel side information. *IEEE Trans. Inform. Theory*, 53(3):881 – 899, Mar. 2007.

[10] J. P. M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback Part I: No bandwidth constraint. *IEEE Trans. Inform. Theory*, IT-12(2):172–182, Apr. 1966.

[11] J. P. M. Schalkwijk. A coding scheme for additive noise channels with feedback Part II:Bandlimited signals. *IEEE Trans. Inform. Theory*, IT-12(2):183–189, Apr. 1966.

[12] N. Elia. When Bode meets Shannon: Control-oriented feedback communication schemes. *IEEE Trans. Automat. Contr.*, 49(9):1477–1488, Sept. 2004.

[13] O. L. V. Costa and M. D. Fragoso. Stability results for discrete-time systems with Markovian jumping parameters. *J. Math. Anal. Applicat.*, 179:154–178, 1993.

[14] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems*. John Wiley & Sons, New York, 1972.

[15] N. Elia and S.K. Mitter. Stabilization of linear systems with limited information. *IEEE Trans. Automat. Contr.*, 46(9):1384–1400, Sept. 2001.

[16] J. Liu and N. Elia. Achieving the stationary feedback capacity for Gaussian channels with memory. *Submitted to IEEE Trans. Inform. Theory*, Oct. 2005. See also: Achieving the Stationary Feedback Capacity for Gaussian Channels. Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing, Sept. 2005. Available at http://arxiv.org/abs/cs.IT/0512097.

[17] S. K. Mitter. Control with limited information. *IEEE Information Theory Society Newsletter*, 50:1–23, Dec. 2000.

[18] S. Tatikonda. *Control Under Communication Constraints*. PhD thesis, MIT, Cambridge, MA, Aug. 2000.

[19] V. S. Borkar, S. K. Mitter, and S. Tatikonda. Markov control problems under communication constraints. *Communications in Information and Systems*, 1:15–32, Jan. 2001.

[20] S. Tatikonda and S. Mitter. The capacity of channels with feedback Part I: the General Case. *Submitted to IEEE Trans. Inform. Theory*, Oct. 2001.

[21] S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Trans. Automat. Contr.*, 49(7):1056 – 1068, July 2004.

[22] S. Tatikonda and S. Mitter. Control over noisy channels. *IEEE Trans. Automat. Contr.*, 49(7):1196 – 1201, July 2004.

[23] A. Sahai. *Anytime Information Theory*. PhD thesis, MIT, Cambridge, MA, 2001.

[24] N. Elia. Stabilization of systems over analog memoryless channels. *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2002.

[25] N. Elia. Feedback stabilization in the presence of fading channels. *Proc. of the 2003 American Control Conference*, pages 4438–4443, June 2003.

[26] A. Sahai and Q. Xu. The anytime reliability of the AWGN+ erasure channel with feedback. *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2004.

[27] J. Liu. *Fundamental Limits in Gaussian Channels with Feedback: Confluence of Communication, Estimation, and Control*. PhD thesis, Iowa State University, Ames, IA, Apr. 2006. Available at http://archives.ece.iastate.edu/archive/00000219/.

[28] J. C. Tiernan and J. P. Schalkwijk. An upper bound on the capacity of the bandlimited Gaussian autoregressive channel with noiseless feedback. *IEEE Trans. Inform. Theory*, IT-20:311–316, 1974.

[29] S. A. Butman. Linear feedback rate bounds for regressive channels. *IEEE Trans. Inform. Theory*, IT-22:363–366, 1976.

[30] L. H. Ozarow. Upper bounds on the capacity of Gaussian channels with feedback. *IEEE Trans. Inform. Theory*, 36(1):156–161, Jan. 1990.

[31] L. H. Ozarow. Random coding for additive Gaussian channels with feedback. *IEEE Trans. Inform. Theory*, 36(1):17–22, Jan. 1988.

[32] A. Shahar-Doron and M. Feder. On a capacity achieving scheme for the colored Gaussian channel with feedback. *Proc. 2004 IEEE International Symposium on Information Theory (ISIT)*, page 74, July 2004.

[33] Young-Han Kim. The feedback capacity of the first-order moving average Gaussian channel. *Accepted by IEEE Trans. Inform. Theory*, 2006. Available at http://arxiv.org/abs/cs/0411036.

[34] N. Merhav and T. Weissman. Coding for the feedback Gel'fand-Pinsker channel and the feedforward Wyner-Ziv source. *IEEE Trans. Inform. Theory*, 52(9):42074211, Sept. 2006.

[35] A. Rosenzweig. The capacity of Gaussian multi-user channels with state and feedback. *IEEE Trans. Inform. Theory*, 53(11):4349–4355, Nov. 2007.

[36] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.

[37] O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, 1997.

[38] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, Nov. 1999.

[39] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley and Sons, 1968.

[40] M. Gastpar. On remote sources and channels with feedback. *Proc. 38th Annual Conference on Information Sciences and Systems (CISS)*, 2004.

[41] J. H. Braslavsky, M. M. Seron, and P. V. Kokotovic. Near-optimal cheap control of nonlinear systems. Technical Report 97-1117, CCEC, University of California, Santa Barbara, CA., 1997.

[42] M. M. Seron, J. H. Braslavsky, P. V. Kokotovic, and D. Q. Mayne. Feedback limitations in nonlinear systems: From bode integrals to cheap control. *IEEE Trans. Automat. Contr.*, 44(4):829–833, Apr. 1999.

[43] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

[44] A. Wyner. On the Schalkwijk-Kailath coding scheme with a peak energy constraint. *IEEE Trans. Inform. Theory*, 14(1):129 – 134, Jan. 1968.

[45] A. Sahai, S. Avestimehr, and P. Minero. Anytime communication over the Gilbert-Eliot channel with noiseless feedback. *Proc. IEEE International Symposium on Information Theory (ISIT)*, Sept. 2005.

[46] E. Ratzer. Low-density parity-check codes on Markov channels. *Presented at the Second IMA Conference on Mathematics in Communications*, 2002. http://www.inference.phy.cam.ac.uk/ear23/papers.

[47] T. E. Klein. *Capacity of Gaussian Noise Channels with Side Information and Feedback*. PhD thesis, MIT, Cambridge, MA, Feb. 2001.

[48] A.P.T. Lau and F. R. Kschischang. Feedback quantization strategies for multiuser diversity systems. *IEEE Trans. Inform. Theory*, 53(4):1386 – 1400, Apr. 2007.

[49] V.K.N. Lau and T. Wu. Optimal transmission and limited feedback design for ofdm/mimo systems in frequency selective block fading channels. *IEEE Trans. Wireless Commun*, 6(5):1569 – 1573, May 2007.

[50] J. Chen, R. A. Berry, and M. L. Honig. Performance of limited feedback schemes for downlink OFDMA with finite coherence time. *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 2751–2755, June 2007.

[51] J. Liu and N. Elia. Writing on dirty paper with feedback. *Communications in Information and Systems*, 5(4):401–422, 2005.

[52] S. M. Kay. *Fundamentals of Statistical Signal Processing II: Detection Theory*. Prentice-Hall PTR, Englewood Cliffs, N.J., 1998.